

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 11:20:07 : Search time 6876 Seconds
(without alignments)
16902.887 Million cell updates/sec

Title: u66617

Perfect score: 2841

Sequence: 1 GAATTCGCGCTATCCCATAG.....TGGGACAAAGGGGATTCC 2841

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Geneml1:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_by:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2841	100.0	2841	9	HSU66617
2	2568.4	90.4	3335	9	BC009368 Homo sapi
3	2427.2	85.4	2746	9	AF109733 Homo sapi
4	1330.4	46.8	3103	6	AX305314 Sequence
5	1330.4	46.8	3103	10	MMU66620 Mus musculus
6	1319.4	46.4	3101	10	MUSDISKZL
7	1308.6	46.1	3086	10	BC026783
8	1246.4	43.9	164813	9	AC025154
9	773	27.2	2557	5	BC045009
10	698.4	24.6	2747	5	BC049347
11	537.6	18.9	1413	12	BT007694
12	537.6	18.9	1698	9	BC002628
13	530.8	18.7	1790	10	BC013122
14	522.8	18.4	1724	9	HSU66619
15	488.8	17.2	606	6	AX395796
16	430	15.1	2379	10	BC005732
17	429	15.1	2297	10	AB003504
18	414.2	14.6	2018	3	AF071503
19	414.2	14.6	2026	3	AY069383
20	414.2	14.6	33444	2	AC014246
21	414.2	14.6	171751	3	AC023697
22	414.2	14.6	174157	3	AC023711
23	414.2	14.6	332029	3	AB003491
24	412.4	14.5	2041	9	HSU66618
25	394	13.9	1674	3	AK112677
26	323.8	11.4	137956	2	AC010918
27	294	10.3	153863	2	AC025361
28	288.2	10.1	110000	2	AC117865_1
29	288.2	10.1	223435	2	AC129640
30	288.2	10.1	237661	2	AC114454
31	284.8	10.0	195301	2	AC134548
32	238.6	8.4	227255	2	AC139317
33	230	8.1	44171	3	AF000265
34	230	8.1	299719	2	AC006858
35	211.8	7.5	294136	2	AC006901
36	203.2	7.2	1060	10	BC012698
37	191.4	6.7	221111	2	AC116568
38	188	6.6	5938	6	AX344810
39	177.4	6.2	236045	2	AC114659
40	177.4	6.2	260277	2	AC114063
41	176.2	6.2	27832	3	CEZK1128
42	168.8	5.9	4006	9	HUMGPD1A
43	145	5.1	5938	6	AX344811
44	123.4	4.3	310	6	AX098486
45	117.4	4.1	160431	8	OSJN00176

ALIGNMENTS

RESULT 1
HSU66617
LOCUS
DEFINITION Human SWI/SNF complex 60 kDa subunit (BAF60a) mRNA, alternatively spliced, complete cds.
ACCESSION U66617
VERSION U66617.1 GI:1549242
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Wang, W., Cole, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R., Murchart, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L. and

Crabtree,G.R.
Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
EMBO J. 15 (1996) In press
2 (bases 1 to 2841)
Wang,W., Xue,Y., Zhou,S., Kuo,A., Cairns,B.R. and Crabtree,G.R.
Diversity and specialization of mammalian SWI/SNF complexes
Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE 96397413
PUBMED 8804307
REFERENCE 3 (bases 1 to 2841)
Wang,W., Xue,Y., Zhou,S. and Crabtree,G.R.
Direct Submission
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
location/Qualifiers
1..2841
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="Jurkat T cells"
1..2841
/gene="BAF60a"
266..1573
/note="alternatively spliced product; present within the 2 MD BRG1 complex; ubiquitously expressed; similar to human BAF60a, BAF60b, BAF60c and yeast SWP73 and SWP73b"
/codon_start=1
/product="SWI/SNF complex 60 kDa subunit"
/protein_id="AAC50695.1"
/db_xref="GI:1549243"
/translation="MGAPAGQLVSRPMPGAAVPRPGMLPGSRMTPOGSPMPGPGYGG NPSVRPLAQAGMDSDRKPAPQIQVQQAQVONRNHAKKXWAKLQRLRELVPESQAMOLLAERKLDITIMRKLDIOEALRKPIKQARKURIFISNFPKSDAED GEGTVASWELRVEGLDLSALSKYDQTKRKFSFSLVILDKDLYGPDNLHVE WHRTATQETDGFQVRPGDVNRCTVLLMDLYOPPOPKLDRPLARLLGIHTQRPVI IQALQYIKTKHLQDPPHEREFVICDKYLOQIFETORMKFSIPORLHALLMPPEPII NHVLSVDNDQKTYACYPDIVEVDLTKTQNSFLSTASQEIATLDNKTMTDVGNG PEERRAEFYQPHQAQEAFCYFYSKVQRRQLEQALGIRNT"
BASE COUNT 621 a 810 c 746 g 564 t
ORIGIN
Query Match 100.0%; Score 2841; DB 9; Length 2841;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2841; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCGGCTATCCCATAGTCTCGCTGCCCTGAGCCCTCCGTCGCCCGCGCGCGCGG 60
DB 1 GAATTCGGCTATCCCATAGTCTCGCTGCCCTGAGCCCTCCGTCGCCCGCGCGCGCGG 60
QY 61 GGGAAACAGCGGGCGCGCGGGGGCGCTCGGGGGGGGGGGAGTTCCCGGTTCCGGTTCTT 120
DB 61 GGGAAACAGCGGGCGCGGGGGCGCTCGGGGGGGGGGGAGTTCCCGGTTCCGGTTCTT 120
QY 121 TGTGGCGGTGCATCGCGGGCTCCGGGAAGATGGCGGGCGGGGGTTCCAGTCTGTGG 180
DB 121 TGTGGCGGTGCATCGCGGGCTCCGGGAAGATGGCGGGCGGGGGTTCCAGTCTGTGG 180
QY 181 CTCGAAGCGGGCGCGCGAGCTCAGGAGGGGGCGGGCGGGCTGCTGCTTGGGCCCGG 240
DB 181 CTCGAAGCGGGCGCGCGAGCTCAGGAGGGGGCGGGCGGGCTGCTGCTTGGGCCCGG 240
QY 241 CGGAATCGGGGGCTCTGTGCGAATGGGGCGGGCGGGCTCGGGTCAAGGGCTGTACCGGTC 300
DB 241 CGGAATCGGGGGCTCTGTGCGAATGGGGCGGGCGGGCTCGGGTCAAGGGCTGTACCGGTC 300
QY 301 CCCGATGCCGAGCGGGCTATCCGAGACCAAGGTATGTTGCCAGCGACCGAATGACACC 360
DB 301 CCCGATGCCGAGCGGGCTATCCGAGACCAAGGTATGTTGCCAGCGACCGAATGACACC 360
QY 361 TCAGGACCTTCCATGGGACCCCTGGCTATGGGGGGAAACCTTCAGTCCGACCTGGCCT 420
DB 361 TCAGGACCTTCCATGGGACCCCTGGCTATGGGGGGAAACCTTCAGTCCGACCTGGCCT 420

361 TCAGGACCTTCCATGGGACCCCTGGCTATGGGGGAAACCTTCAGTCCGACCTGGCCT 420
QY 421 GGCCCAAGTCAGGGATGGATCAGTCCCGCAAGAGACCTTGCCTCAGCAGATCCAGCAGGT 480
DB 421 GGCCCAAGTCAGGGATGGATCAGTCCCGCAAGAGACCTTGCCTCAGCAGATCCAGCAGGT 480
QY 481 CCAGCAGCGCGGTCCAAATCGAAACCAATGCAAAAGAAAAAGAAAGATGGCTGACAA 540
DB 481 CCAGCAGCGCGGTCCAAATCGAAACCAATGCAAAAGAAAAAGAAAGATGGCTGACAA 540
QY 541 AATTCTACCTCAAGAGGATTCGTGAACCTGGTACCAAGATCCAGGCTTATATGGATCTCTT 600
DB 541 AATTCTACCTCAAGAGGATTCGTGAACCTGGTACCAAGATCCAGGCTTATATGGATCTCTT 600
QY 601 GGCTTTTGAAGAGAAAAGCTGGACCAAGATATCATGAGGAAACGGCTAGATATCCAAAGAGGC 660
DB 601 GGCTTTTGAAGAGAAAAGCTGGACCAAGATATCATGAGGAAACGGCTAGATATCCAAAGAGGC 660
QY 661 CTTGAAAAGCTCCATTAAAGCAAAAACGGAAGCTCGAATTTTCAATTTCTAACTTTCAA 720
DB 661 CTTGAAAAGCTCCATTAAAGCAAAAACGGAAGCTCGAATTTTCAATTTCTAACTTTCAA 720
QY 721 TCCGGCTAAGTCAGATCCCGAGGATGGGGAAGGACCGTGGCTTCTCGGAGCTTCGGGT 780
DB 721 TCCGGCTAAGTCAGATCCCGAGGATGGGGAAGGACCGTGGCTTCTCGGAGCTTCGGGT 780
QY 781 AGAAGGACGGCTCTCGAGGATTCAGCCTTGTCCAAATATGATGCCACTAAACAAAAGAG 840
DB 781 AGAAGGACGGCTCTCGAGGATTCAGCCTTGTCCAAATATGATGCCACTAAACAAAAGAG 840
QY 841 GAACTTCTCTTCTTTTAAAGTCTTGGTGAATGAACTGGCAAAAGACCTGTATGGGCC 900
DB 841 GAACTTCTCTTCTTTTAAAGTCTTGGTGAATGAACTGGCAAAAGACCTGTATGGGCC 900
QY 901 AGACAAACATCTGTAGATGCGACAGACCGCCATACCCAGAGACCGATGGCTTCCA 960
DB 901 AGACAAACATCTGTAGATGCGACAGACCGCCATACCCAGAGACCGATGGCTTCCA 960
QY 961 GGTCAAGCGGCGAGAGATGTAATGTACGGTGTACTGTCTCTATGTGCTGGATTACCA 1020
DB 961 GGTCAAGCGGCGAGAGATGTAATGTACGGTGTACTGTCTCTATGTGCTGGATTACCA 1020
QY 1021 GCCTCCCGCAGTTTAAATTAGACCCCGCCTAGCTCGACTCCTGGGCATCCATACCCAGAC 1080
DB 1021 GCCTCCCGCAGTTTAAATTAGACCCCGCCTAGCTCGACTCCTGGGCATCCATACCCAGAC 1080
QY 1081 TCCTCAGTGTATCATCAAGCAGTGTGGCAATATATTAAGACACATAGCTCCAGGACCC 1140
DB 1081 TCCTCAGTGTATCATCAAGCAGTGTGGCAATATATTAAGACACATAGCTCCAGGACCC 1140
QY 1141 TCAGAGCGGGAGTTTGTCTATCTGTGACAAAGTACCTCGCAGCAGATCTTTGAGACTCAACG 1200
DB 1141 TCAGAGCGGGAGTTTGTCTATCTGTGACAAAGTACCTCGCAGCAGATCTTTGAGACTCAACG 1200
QY 1201 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTATGTCACAGAACCTTAT 1260
DB 1201 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTATGTCACAGAACCTTAT 1260
QY 1261 CATCAATTAATCATGTCTAGTGTGACCCGGAATGATCAGAAAAGACAGCTTGTATGA 1320
DB 1261 CATCAATTAATCATGTCTAGTGTGACCCGGAATGATCAGAAAAGACAGCTTGTATGA 1320
QY 1321 CATTGATGTTGAAGTGGATGACACCTTGAAGACCCAGATGAATTTCTTTCGCTGTCCAC 1380
DB 1321 CATTGATGTTGAAGTGGATGACACCTTGAAGACCCAGATGAATTTCTTTCGCTGTCCAC 1380
QY 1381 TGCCAGCAACAGAGAGATTGCTACTCTAGACAA CAAGACAAATGATGATGTGGGTGATA 1440
DB 1381 TGCCAGCAACAGAGAGATTGCTACTCTAGACAA CAAGACAAATGATGATGTGGGTGATA 1440
QY 1441 CCCAGAGAGAGAGCGCGAGCTGATGTTCTATCTCCAGCCTGGGCTCAGGAGGCTGTGTG 1500
DB 1441 CCCAGAGAGAGAGCGCGAGCTGATGTTCTATCTCCAGCCTGGGCTCAGGAGGCTGTGTG 1500

Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhi.nih.gov

BASE COUNT	746 a	954 c	846 g	789 t	ADJACENT ORIGIN
Query Match	90.4%;	Score 2568.4;	DB 9;	Length 3335;	
Best Local Similarity	95.1%;	Pred. No. 0;			
Matches 2798;	Conservative 0;	Mismatches 11;	Indels 133;	Gaps 2;	
Qy	31	CTGAGCCTCCCGTGC	CGCGCGCGCGCGGAA	CAGGCGGGCGC	-CGGGGGCGGCTCG 89
Db	10	CTGAGCCTCCCGTGC	CGCGCGCGCGGGGAA	CAGGCGGGCGCTCG	GGGGGGCGGCTCG 69
Qy	90	GGGGGGGGGGGAGTTC	CGGTTCCGGTTC	TGTGGCGTGCAT	CGCGCGCTCGGGGAG 149
Db	70	GGGGGGGGGGGAGTTC	CGGTTCCGGTTC	TGTGGCGTGCAT	CGCGCGCTCGGGGAG 129
Qy	150	ATGGCGCCCGGGGGG	TTTCCAGTCTGTGGCT	TCCAAAGCGCGCGCGG	CGGAGCCTCAGGA 209
Db	130	ATGGCGCCCGGGGGG	TTTCCAGTCTGTGGCT	TCCAAAGCGCGCGCGG	CGGAGCCTCAGGA 189
Qy	210	GGGGCGGGCGGCTGT	CGCTTGGGGCC	-GGCGGAATC	CGGGGCTCTCTGTCCGAATG 268
Db	190	GGGGCGGGCGGCTGT	CGCTTGGGGCC	GGGGCGGAATC	CGGGGCTCTCTGTCCGAATG 249
Qy	269	GGGCGCGCTCCGGT	CAAGGGGCTGTAC	CGCTCCCCGATG	CCCGAGCGGCTATCCGAGA 328

QY 1409 GACAA----- 1413
 Db 1390 GACCAACAAATGCATAGAGCAATAGAAACCATCAACAGCTGAAGCTCAGGGGAGTTCC 1449
 QY 1414 ----- 1413
 Db 1450 ATGTAGAGCTTTGCCAGAGACCCCTCAGGGTTTCATCAATGACTGGCTTCAGTCCAGTCC 1509
 QY 1414 ----- 1465
 Db 1510 AGGCACTTCAAGACATATCATATGTGGGGTAAACCAAGAGAGAGCGCCAGACTGAG 1569
 QY 1466 TTCTACTTCCAGCCCTGGGCTCAGAGAGGTGTGTCCGATACCTTCACTTCAAGGTGACG 1525
 Db 1570 TTCTACTTCCAGCCCTGGGCTCAGAGAGGTGTGTCCGATACCTTCACTTCAAGGTGACG 1629
 QY 1526 CAGAGACGACAAAGATTAGAGCAGCCCTGGGAATCCGGAATACATAGGGCTTCCAC 1585
 Db 1630 CAGAGACGACAAAGATTAGAGCAGCCCTGGGAATCCGGAATACATAGGGCTTCCAC 1689
 QY 1586 AGCCCTGATTCGACCTGACCAATTTCTGATTTGGGCCCTGTGCTGCTGCTCATATGAT 1645
 Db 1690 AGCCCTGATTCGACCTGACCAATTTCTGATTTGGGCCCTGTGCTGCTGCTCATATGAT 1749
 QY 1646 CTGGCTTGGTCTTGGCTGGGGGCTTCCAGGGGATGCTGTGTTCAAGGACCAAGACCA 1705
 Db 1750 CTGGCTTGGTCTTGGCTGGGGGCTTCCAGGGGATGCTGTGTTCAAGGACCAAGACCA 1809
 QY 1706 ATGAAGAGGGTCTCAACAGACACTGTTATCTCTCTTCTTCAACCTATCTTCTCCACCC 1765
 Db 1810 ATGAAGAGGGTCTCAACAGACACTGTTATCTCTCTTCTTCAACCTATCTTCTCCACCC 1869
 QY 1766 CAGAGCTTCCCTTGGCCCAAGAGTTCCCATGTGCTGTACCTCCCTGGTCTACATG 1825
 Db 1870 CAGAGCTTCCCTTGGCCCAAGAGTTCCCATGTGCTGTACCTCCCTGGTCTACATG 1929
 QY 1826 GACCTCTAGATAGTGTAGAGAGAACATGATGTTAGTGTGTTAGTGTGTTGATGATG 1884
 Db 1930 GACCTCTAGATAGTGTAGAGAGAACATGATGTTAGTGTGTTAGTGTGTTGATGATG 1989
 QY 1885 GGCCTCAGGCGCAGTGTCTTCAAGGGGACAGCTACTGATCTTCAACCTTCCAGAGCC 1944
 Db 1990 GGCCTCAGGCGCAGTGTCTTCAAGGGGACAGCTACTGATCTTCAACCTTCCAGAGCC 2049
 QY 1945 AGGAGTTGGG---TTTGGCTCTTCTTCCAGACTCAGGCTGTGGGACTCTATTAAGCTA 2001
 Db 2050 AGGAGTTGGGAGCTTCTCTCTCTTCCAGACTCAGGCTGTGGGACTCTATTAAGCTA 2109
 QY 2002 GTTGTATCTTGGCTCTCCGATTAACAGAACTCAATTCCTTCCCTCCCAAGGTTTG 2061
 Db 2110 GTTGTATCTTGGCTCTCCGATTAACAGAACTCAATTCCTTCCCTCCCAAGGTTTG 2169
 QY 2062 AACAACTCTCCCTTCACTTGTGTGCTGTAGCACTACAGAAACCTGTTCTT- GGCTC 2120
 Db 2170 AACAACTCTCCCTTCACTTGTGTGCTGTAGCACTACAGAAACCTGTTCTTGGGCTC 2229
 QY 2121 CACTAGAGCCCGAGTCAAGTCCCGAG- CCTGTGGGTTGGGCTGTGCTGCTGCTCTCA 2179
 Db 2230 CACTAGAGCCCGAGTCAAGTCCCGAGCCTGTGGGTTGGGCTGTGCTGCTCTCA 2289
 QY 2180 CTCCCTAGTGGGGTCCACATAGATAGTGGATTTTGTCTTATGTCTCCCTCCCAAGC 2239
 Db 2290 CTCCCTAGTGGGGTCCACATAGATAGTGGATTTTGTCTTATGTCTCCCTCCCAAGC 2349
 QY 2240 ACTCCCTGTGGCTGCTTGTGATTCCTCAGATCTGCTCAATCCCGGCAATTGGGT 2299
 Db 2350 ACTCCCTGTGGCTGCTTGTGATTCCTCAGATCTGCTCAATCCCGGCAATTGGGT 2409
 QY 2300 GGGGGAAATCTTGGCTTCTTCCCTTCCAGAGCCCGAGGATCTCATCTGGGAACTGTCAATTG 2359
 Db 2410 GGGGGAAATCTTGGCTTCTTCCCTTCCAGAGCCCGAGGATCTCATCTGGGAACTGTCAATTG 2469

QY 2360 CCAGCAGAGGCTGTTCCCTTCTGAGTGTGAGATGTGATCTATTCATTCATCTGCTCC 2419
 Db 2470 CCAGCAGAGGCTGTTCCCTTCTGAGTGTGAGATGTGATCTATTCATTCATCTGCTCC 2528
 QY 2420 ACCCTGCTCTGCACTCCCTTAAATGAGAAAGGGGCTTAAACCAACGGGTTAAAAAGCC 2479
 Db 2529 ACCCTGCTCTGCACTCCCTTAAATGAGAAAGGGGCTTAAACCAACGGGTTAAAAAGCC 2587
 QY 2480 CTGGGCACTCCCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2539
 Db 2588 CTGGGCACTCCCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2647
 QY 2540 AGGCACTGAGAGTGAAGAGGCTTGAAGGCTGAGAAATAGCGCTGATGTGGTTGTGAC 2599
 Db 2648 AGGCACTGAGAGTGAAGAGGCTTGAAGGCTGAGAAATAGCGCTGATGTGGTTGTGAC 2707
 QY 2600 TCTTCCCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2659
 Db 2708 TCTTCCCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2767
 QY 2660 ACCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGGCCCTTGTGATCTGATGTTTGG 2719
 Db 2768 ACCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGGCCCTTGTGATCTGATGTTTGG 2827
 QY 2720 CATCTTCCAGAGAGCCTTCAATCTTCTTCCAGGTTGATACCCCGAGTTAGACATA 2779
 Db 2828 CATCTTCCAGAGAGCCTTCAATCTTCTTCCAGGTTGATACCCCGAGTTAGACATA 2887
 QY 2780 TCCGAGGCTGCGAGCTCAACACAGCAGGTTGGGAGACAGCTGGGCAACAAAGGGGATT 2839
 Db 2888 TCCGAGGCTGCGAGCTCAACACAGCAGGTTGGGAGACAGCTGGGCAACAAAGGGGATT 2947
 QY 2840 CC 2841
 Db 2948 TC 2949

RESULT 3
 AF109733
 LOCUS
 DEFINITION
 Homo sapiens SWI/SNF-related, matrix-associated, actin-dependent
 regulator of chromatin D1 (SMARCD1) mRNA, complete cds.
 ACCESSION
 AF109733
 VERSION
 AF109733.1 GI:4566529
 KEYWORDS
 ORGANISM
 Homo sapiens (human)
 SOURCE
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2746)
 Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.
 Diversity and specialization of mammalian SWI/SNF complexes
 Genes Dev. 10 (17), 2117-2130 (1996)
 JOURNAL
 MEDLINE
 PUBMED
 9639743
 8804307
 2 (bases 1 to 2746)
 Ring, H. Z., Vameghi-Meyers, V., Wang, W., Crabtree, G. R. and Francke, U.
 Five SWI/SNF-related, matrix-associated, actin-dependent regulator
 of chromatin (SMARC) genes are dispersed in the human genome
 Genomics 51 (1), 140-143 (1998)
 JOURNAL
 MEDLINE
 PUBMED
 9639744
 3 (bases 1 to 2746)
 Ring, H. Z. and Francke, U.
 Direct Submission
 Submitted (25-NOV-1998) Genetics and HHMI, Stanford University
 School of Medicine, B205 Beckman Center, Stanford, CA 94305, USA
 FEATURES
 source
 1..2746
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="12"

```

/map="12q13-q14"
/tissue type="peripheral blood"
/note="From two individuals"
1..2746
/gene="SMARCD1"
145..1506
/gene="SMARCD1"
/note="BAF60a; similar to sequences presented in GenBank
Accession Numbers U66617, U66618, U66619 and U66620"
/codon_start=1
/product="SWI/SNF-related, matrix-associated,
actin-dependent regulator of chromatin D1"
/protein_id="AAD23390.1"
/db_xref="GI:4566530"
/translacion="MLPSRMTPOGSPMGPGYGNPSVRPLAOSGMDQSKRPAPQ
OIQQVQQAQVQRNHNNAKKWADKILPORIRELPESQAYMDLLAFERKLDQIMRK
RLDQEALEKRFIKQKRKURIFISNTFPAKSDAEDGEGTVASWELRVEGRLLSALS
KYDQAKRKSSFFKSVIELDXYLDYGNHLEWHRTATQETDGFQVRRPGGVNV
RCTLVLLDYQPPQKLPRLARLGIHTQRPVLIQALWQYIKTHKLDQPHREFVI
CDKYLQOIFESORMKFSIPORLHALLPPEPIIINHVISVDPNDKKIACYDIDVEI
DRTLKTONSFLLSASQOEIATLQNKIHTIETINOLKTQREPLSPARPOQPFIND
WLOSQRDLKMTWDVVGNEPEERRAEFFYPQWAEAVCRYFYFVKVQQRQLEQALGI
RNT"
misc_difference 1228..1350
/gene="SMARCD1"
/note="difference between this sequence and Homo sapiens
sequence presented in GenBank Accession Number U66617"
/replace="-"

BASE COUNT      628 a  777 c  686 g  655 t
ORIGIN

Query Match      85.4%; Score 2427.2; DB 9; Length 2746;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 2611; Conservative 0; Mismatches 8; Indels 129; Gaps 5;

QY 192 GCGCGGAGCTCAGGAGGCGGCGCGGTGCTGCTTGGGCC-GGCGGAATCGG 250
DB 1 GCGCGGAGCTCAGGAGGCGGCGCGGTGCTGCTTGGGCCGGGAACTCGG 60
QY 251 GGGCTCTGTGGAATGGCGCGGCTCGGGTCAAGGCTACCGCTCCCGATGCC 310
DB 61 GGGCTCTGTGGAATGGCGCGGCTCGGGTCAAGGCTACCGCTCCCGATGCC 120
QY 311 GGAGCGGCTATCCGAGACCAAGTATGTTGCCAGGCGCGCAATGACACCTCAGGAGCT 370
DB 121 GGAGCGGCTATCCGAGACCAAGTATGTTGCCAGGCGCGCAATGACACCTCAGGAGCT 180
QY 371 TCCATGGGACCCCTGGCTATCGGGGAA CTTTCAGTCCGACCTGGGCTGGCCCACTCA 430
DB 181 TCCATGGGACCCCTGGCTATCGGGGAA CTTTCAGTCCGACCTGGGCTGGCCCACTCA 240
QY 431 GGGATGGATCAGTCCCGAAGAGACTGCCCTCAGCAGATCCAGCAGGTCCAGCAGAG 490
DB 241 GGGATGGATCAGTCCCGAAGAGACTGCCCTCAGCAGATCCAGCAGGTCCAGCAGAG 300
QY 491 GGGTCCAAAATCGAAA CCAATGCAAGAAAAGAGATGGCTGACAAAATTTCTACT 550
DB 301 GGGTCCAAAATCGAAA CCAATGCAAGAAAAGAGATGGCTGACAAAATTTCTACT 360
QY 551 CAAAGGATTCGTGAA CTGTTACAGAA TCCAGGCTTATATGATCTCTTGGCTTTTGA 610
DB 361 CAAAGGATTCGTGAA CTGTTACAGAA TCCAGGCTTATATGATCTCTTGGCTTTTGA 420
QY 611 AGGAACTGGACCACTATCATAGGAAACCGCTAGATATCAAGAGGCTTGAACGT 670
DB 421 AGGAACTGGACCACTATCATAGGAAACCGCTAGATATCAAGAGGCTTGAACGT 480
QY 671 CCATTAAACAAAACGGAAGCTCGAA TTTTCATTCTTAAACATTTTCAATCCGGCTAAG 730
DB 481 CCATTAAACAAAACGGAAGCTCGAA TTTTCATTCTTAAACATTTTCAATCCGGCTAAG 540
QY 731 TCAGATCGCGAGGATGGGAAAGGAGCGGTGGCTTCTGGGAGCTTCGGGTAGAAAGCGG 790
DB 121 TCAGATCGCGAGGATGGGAAAGGAGCGGTGGCTTCTGGGAGCTTCGGGTAGAAAGCGG 600
QY 791 CTCCTGGAGGATTCAGCCTTGTCCAAATATATGATGCACTAAACAAAAGAGAAATTTCTCT 850
DB 601 CTCCTGGAGGATTCAGCCTTGTCCAAATATATGATGCACTAAACAAAAGAGAAATTTCTCT 660
QY 851 TCCCTTTTAAAGTCTTGGTGATTTGAACTGACAAAAGACCTGTATGGGCCAGAACCAAT 910
DB 661 TCCCTTTTAAAGTCTTGGTGATTTGAACTGACAAAAGACCTGTATGGGCCAGAACCAAT 720
QY 911 CTGCTAGAAATGGCAGAGACCCCACTACCCAGAGACCGATGGCTTCCAGGTGAAGCGG 970
DB 721 CTGCTAGAAATGGCAGAGACCCCACTACCCAGAGACCGATGGCTTCCAGGTGAAGCGG 780
QY 971 CCAGAGATGTGAATGTACGGTGTACTGTCTCTACTGTCTGGATTAACAGCCTCCCCAG 1030
DB 781 CCGGAGACGTGAATGTACGGTGTACTGTCTCTACTGTCTGGATTAACAGCCTCCCCAG 840
QY 1031 TTTAAATTTAGACCCCGCTAGCTCGACTCTCTGGGCATCCATAGCTCGTCCAGTG 1090
DB 841 TTTAAATTTAGACCCCGCTAGCTCGACTCTCTGGGCATCCATAGCTCGTCCAGTG 900
QY 1091 ATCATCCAGCACCTGTGGCAATATATTAAGACACATAGCTCCAGGACCTCCAGAGCGG 1150
DB 901 ATCATCCAGCACCTGTGGCAATATATTAAGACACATAGCTCCAGGACCTCCAGAGCGG 960
QY 1151 GAGTTTGTCTCTGTGACAGTACCTGACAGAGATCTTTGAGACTCAACGCTATGAAGTTT 1210
DB 961 GAGTTTGTCTCTGTGACAGTACCTGACAGAGATCTTTGAGACTCAACGCTATGAAGTTT 1020
QY 1211 TCAGAGATTCCTCAGCGGCTCCATGCTTGTATGCAACAGAACTTATCATCATTAAT 1270
DB 1021 TCAGAGATTCCTCAGCGGCTCCATGCTTGTATGCCACAGAACTTATCATCATTAAT 1080
QY 1271 CATGTCATCAGTGTGACCCGAATGATCAGAAAAGAGAGCTTGTATGACATGATGTT 1330
DB 1081 CATGTCATCAGTGTGACCCGAATGATCAGAAAAGAGAGCTTGTATGACATGATGTT 1140
QY 1331 GAGTTGGATGACACTTGAAGACCCAGATGAATTTCTTTCTGCTCCACTGCCAGCA 1390
DB 1141 GAGTTGGATGACACTTGAAGACCCAGATGAATTTCTTTCTGCTCCACTGCCAGCA 1200
QY 1391 CAGGAGATTTGCTACTCTAGACAA ----- 1413
DB 1201 CAGGAGATTTGCTACTCTAGACAA CAAAGATCCATGAGACAA TAGAAA CCAATCAACAGCTG 1260
QY 1414 ----- 1413
DB 1261 AAGACTCAGCGGAGTTTCACTGAGCTTTGCCAGAGACCTTCAGGGTTTTCATCAATGAC 1320
QY 1414 ----- CAAAGCAATGACTGATGTGGTGGTAA CCGAGAG 1447
DB 1321 TGGCTTCACTCCAGTGCAGGAGCTCAAGACAAATGACTGATGTGGTGGTAA CCGAGAG 1380
QY 1448 GAGGAGCGCCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGGAGGCTGTGTGCCGATAC 1507
DB 1381 GAGGAGCGCCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGGAGGCTGTGTGCCGATAC 1440
QY 1508 TTTACTCAAGGTGTCAGCAGAGAGCAAGAAATTAGAGAGCCCTGGGAATCCGGAAT 1567
DB 1441 TTTACTCAAGGTGTCAGCAGAGAGCAAGAAATTAGAGAGCCCTGGGAATCCGGAAT 1500
QY 1568 ACATAGGCGCTCTCCACAGCCCTGATTCGACTGCACCAATTTCTTGATTTGGGCCCTGTG 1627
DB 1501 ACATAGGCGCTCTCCACAGCCCTGATTCGACTGCACCAATTTCTTGATTTGGGCCCTGTG 1560
QY 1628 CTGCTTGCCTCATAGTATCTGCTTGGTCTTGGTGGCGGTTCCAGGGGATGCTGTTGG 1687
DB 1561 CTGCTTGCCTCATAGTATCTGCTTGGTCTTGGTGGCGGTTCCAGGGGATGCTGTTGG 1620
QY 1688 TTCAAGGACAGACCAAGAAATGAAGGGGTCTCACAAGACACCTGTTATCTCTCTTTCA 1747
DB 1621 TTCAAGGACAA CACCAAGAAATGAAGGGGTCTCACAAGACACCTGTTATCTCTCTTTCA 1680
```

1748 CCTATCTCTCCACCCCCAGCTTCCCTTGGCCCCCAAAAGTCCCATGTCCTGTACC 1807
1681 CCTATCTCTCCACCCCCAGCTTCCCTTGGCCCCCAAAAGTCCCATGTCCTGTACC 1740
1808 CTCCCTGGCTACATAGGACCTCTAGATAGTGTAGAGAGAGAA CATGTAGTGTAAATG 1867
1741 CTCCCTGGCTACATAGGACCTCTAGATAGTGTAGAGAGAGAA CATGTAGTGTAAATG 1800
1868 AGTGTGTGAATGGAATTTGGCCCTCAGGCCAGGTGTCTTCAAGGGACCCAGCTAACTGATC 1927
1801 AGTGTGTGAATGGAATTTGGCCCTCAGGCCAGGTGTCTTCAAGGGACCCAGCTAACTGATC 1860
1928 CTACCTCTCAGAGACCCAGAGAGTGGG--TTTGGCTCCCTTCCAAAGCTCAGGCTGT 1984
1861 CTACCTCTCAGAGACCCAGAGAGTGGGAGCTTCCGCTCTTCCAAAGCTCAGGCTGT 1920
1985 GGGCACTGTATAGGTAGTGTGATCTTGGCTCTCTGATTAACAGAAATCCAAATTTCTTCT 2044
1921 GGGCACTGTATAGGTAGTGTGATCTTGGCTCTCTGATTAACAGAAATCCAAATTTCTTCT 1980
2045 TCCCTCCAGAGTGTGGAAACAACTCTCCCTCACTTGTGGCCCTGTAGACTACAGAAA 2104
1981 TCCCTCCAGAGTGTGGAAACAACTCTCCCTCACTTGTGGCCCTGTAGACTACAGAAA 2040
2105 CCCTGGTCTTGGCTCCACTGAGCCCGAGGTGAGTCCCGAGCTCGGGTGGCTGGCTG 2164
2041 CCCTGGTCTTGGCTCCACTGAGCCCGAGGTGAGTCCCGAGCTCGGGTGGCTGGCTG 2100
2165 TCAGTGTCTCTCACTCTCTAGTGTGGGTCCACATCAGTATTTGATTTGTTCTTAT 2224
2101 TCAGTGTCTCTCACTCTCTAGTGTGGGTCCACATCAGTATTTGATTTGTTCTTAT 2160
2225 TGCTCCCTCCAGACACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2284
2161 TGCTCCCTCCAGACACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
2285 CCGGGGCACTTGGGGGGGGGGAATCTTGGCTTCCCTTTCAGAGCCCGAGGATCTCATCT 2344
2221 CCGGGGCACTTGGGGGGGGGGAATCTTGGCTTCCCTTTCAGAGCCCGAGGATCTCATCT 2280
2345 GGGGAATGTCACTTCCAGACAGAGCTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2404
2281 GGGGAATGTCACTTCCAGACAGAGCTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
2405 CCATTCACCTCACTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2464
2341 -CATTCACCTCACTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2399
2465 AGGGGTAAAAAAGCCCTGGGCGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2524
2400 AGGGGT-AAAAAGCCCTGGGCGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2458
2525 CTACTGTGACTTCTAGAGGCACTGAGAGTGAAGCGCCTAGGAGCTGAGAAATAGCGCTG 2584
2459 CTACTGTGACTTCTAGAGGCACTGAGAGTGAAGCGCCTAGGAGCTGAGAAATAGCGCTG 2518
2585 AGTTGGGTTTGTGACTTCTTCCCTCTCTCTGCTGCTCAAGAGATTGATCTCCAGGCCCTG 2644
2519 AGTTGGGTTTGTGACTTCTTCCCTCTCTCTGCTGCTCAAGAGATTGATCTCCAGGCCCTG 2578
2645 CCCTAAAGCTTCAACACCCCTCAGGTAGAGAGAGACCTTGTGATCTTGGCCCTTGGAT 2704
2579 CCCTAAAGCTTCAACACCCCTCAGGTAGAGAGAGACCTTGTGATCTTGGCCCTTGGAT 2638
2705 CTGAGATGGTCTTTCATCTTTCAGAGAGAGCTCACTTCTTCTCCAGTGTGATCAC 2764
2639 CTGAGATGGTCTTTCATCTTTCAGAGAGAGCTCACTTCTTCTCCAGTGTGATCAC 2698
2765 CCCCAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGGTG 2812
2639 CCCCAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGGTG 2746

RESULT 4
AX305314 3103 bp DNA linear PAT 11-DEC-2001
LOCUS
DEFINITION Sequence 65 from Patent WO0188188.
ACCESSION AX305314
VERSION AX305314.1 GI:17644884
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurignathi; Muridae; Murinae; Mus.
REFERENCE
1 Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.
AUTHORS
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 65 22-NOV-2001;
School Uniridical Person Nihon University (JP)
FEATURES
source
location/Qualifiers
1..3103
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
BASE COUNT 657 a 925 c 817 g 704 t
ORIGIN
Query Match 46.8%; Score 1330.4; DB 6; Length 3103;
Best Local Similarity 75.3%; Pred. No. 0;
Matches 2124; Conservative 0; Mismatches 431; Indels 265; Gaps 27;
158 CCGGGCGGGTTTCCAGTGTGCTTCAAGCGCGCGCGAGGCTCAGAGAGGGCGGG 217
1 CCGGGCGGGTTTCCAGTGTGCTTCAAGCGCGCGCGAGGCTCAGAGAGGGCGGG 60
218 CGCGGCTCTGCTTGGGCGCGGAACTCGGGGCTCTCTGTGCGAATGGGCCCGCT 277
61 CGTGGCGGCTCTCTGCGCGCGCGGAACTCGGGGCTCTCTGTGCGAATGGGCCCG 120
278 CCGGGTCAAGGCGGTACCGGCTCCCGGATGCGGGGCGCTTCCAGACAGAGATG 337
121 CCGGGTCAAGGCGGTACCGGCTCTCGATGCGGGGCGCTTCCAGACAGAGATG 180
338 TTGCGAGGACCGGATGACACTCAAGGACCTTTCATGAGACCCCTGTATGGGGG 397
181 CTGCGAGGTACCGGATGACACTCAAGGACCTTTCATGAGACCCCTGTATGGGGG 240
398 AACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 457
241 AACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
458 GCCCTGACGATTCAGACAGTCCAGAGAGCGGTCGCAAAATCGAAATCGAAATCGA 517
301 GCACTTCAACGATTCAGACAGTCCAGAGAGCGGTCGCAAAATCGAAATCGAAATCGA 360
518 AAGAAAGAGATGCTGACAAATTTACTTCAAGAGATTCTGTAAGTGGTACAGAA 577
361 AAGAAAGAGATGCTGACAAATTTACTTCAAGAGATTCTGTAAGTGGTACAGAA 420
578 TCCAGGCTTATGATGATCTTGGCTTTTGAAGAAATGAGACAGACTATCATGAGG 637
421 TCACAGGCTTATGATGATCTTGGCTTTTGAAGAAATGAGACAGACTATCATGAGG 480
638 AAAGGCTAGATTCAGAGAGGCTTGAAGTCCCATTAAGCAAAACGAAAGCTGGA 697
481 AAAGGCTAGATTCAGAGAGGCTTGAAGTCCCATTAAGCAAAACGAAAGCTGGA 540
698 ATTTTCATTTTCAACACTTCAATCCGCTAAGTCAGATGCCAGATGGGAAAGGAG 757
541 ATTTTCATTTTCAACACTTCAATCCGCTAAGTCAGATGCCAGATGGGAAAGGAG 600
758 GTGGCTTCTGGAGCTTGGGTAGAAAGAGCGCTCTGAGGATTCAGCTTGTCAA 817
601 GTGGCTTCTGGAGCTTGGGTAGAAAGAGCGGCTCTGAGGAGAGCGGCTTGTCAA 660


```

MMU66620      3103 bp  mRNA  linear  ROD 18-SEP-1996
LOCUS      Mus musculus SWI/SNF complex 60 KDa subunit (BAF60a) mRNA, complete
DEFINITION
ACCESSION   U66620
VERSION     U66620.1 GI:1549248
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM
REFERENCE   1 (bases 1 to 3103)
AUTHORS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R.,
            Muchardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L. and
            Crabtree, G.R.
            Purification and biochemical heterogeneity of the mammalian SWI-SNF
            complex
            EMBO J. 15 (1996) In press
REFERENCE   2 (bases 1 to 3103)
AUTHORS     Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R. and Crabtree, G.R.
TITLE       Diversity and specialization of mammalian SWI/SNF complexes
JOURNAL     Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE    96397413
PUBMED     8804307
REFERENCE   3 (bases 1 to 3103)
AUTHORS     Wang, W., Xue, Y., Zhou, S. and Crabtree, G.R.
TITLE       Direct Submission
JOURNAL     Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford
            University, Beckman Center B207, Stanford, CA 94305-5428, USA
FEATURES
            source
            1..3103
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /note="similar to musd1sk21, GenBank Accession Number
                1..3103
                /gene="BAF60a"
                /db_xref="GI:1549249"
                /note="ubiquitously expressed; similar to human BAF60a,
                BAF60b, BAF60c and yeast Swp73 and Swp73b"
                /codon_start=1
                /product="SWI/SNF complex 60 KDa subunit"
                /protein_id="AAC52794.1"
                /translation="MGPAFGQLYRSPMPGAAYPRPGLPGSRMTPOGSPMPGPGY
                GPSVPGLAOSGMDOSRRKRPAPQIOYQOQAVONRNNAKKKKADKILPQRIELV
                PESQVYMLDAFERKLDOTIMRKLDIEALRPIKORKKRIFISNFPNPKDAED
                GEGTVASWELRVEGRLLLEDALSKYDARKKRSSEFKSLVIELDKLXLPEDNHLVE
                WHRTATGETDFQVYKRPEDVAVRCTVILMDYOPQPKLDRRLKLGHTQRPVI
                IQALWQYIKTHRLQDPHERFVLCQKYLQQLFESQRMKFSLEIPRLHALLMPEPILII
                NHVLSVDNDQKTKACDYDIVEVDLTKTQNSFLSTASQOEIATLNNKIHETIETI
                NOLKTQREFMLSPESGFINMLQSOGRGLKTMVDVGNPEERAEAFYFQPMWQEA
                VCRYSVYQGRQRELQALGIRMT"
            CDS
                gene
                109..1536
                /gene="BAF60a"
                /note="ubiquitously expressed; similar to human BAF60a,
                BAF60b, BAF60c and yeast Swp73 and Swp73b"
                /codon_start=1
                /product="SWI/SNF complex 60 KDa subunit"
                /protein_id="AAC52794.1"
                /translation="MGPAFGQLYRSPMPGAAYPRPGLPGSRMTPOGSPMPGPGY
                GPSVPGLAOSGMDOSRRKRPAPQIOYQOQAVONRNNAKKKKADKILPQRIELV
                PESQVYMLDAFERKLDOTIMRKLDIEALRPIKORKKRIFISNFPNPKDAED
                GEGTVASWELRVEGRLLLEDALSKYDARKKRSSEFKSLVIELDKLXLPEDNHLVE
                WHRTATGETDFQVYKRPEDVAVRCTVILMDYOPQPKLDRRLKLGHTQRPVI
                IQALWQYIKTHRLQDPHERFVLCQKYLQQLFESQRMKFSLEIPRLHALLMPEPILII
                NHVLSVDNDQKTKACDYDIVEVDLTKTQNSFLSTASQOEIATLNNKIHETIETI
                NOLKTQREFMLSPESGFINMLQSOGRGLKTMVDVGNPEERAEAFYFQPMWQEA
                VCRYSVYQGRQRELQALGIRMT"
BASE COUNT      657 a      925 c      817 g      704 t
ORIGIN
Query Match      46.8%; Score 1330.4; DB 10; Length 3103;
Best Local Similarity 75.3%; Pred. No. 0;
Matches 2124; Conservative 0; Mismatches 431; Indels 265; Gaps 27;
QY      CCGGGCGGGTTTCCAGTCTGTGGCTCCAAAGCGGCGCGGAGCTTCAGAGGGGCGGG 217
DB      1 CCGGGCGGGTTTCCAGTCTGTGGCTCCGAGCGGTGCGGAGGCTTCAGAGGGGCGGG 60
QY      CGCGGCTGTGCTTGGGCGCGGCGGAATTCGGGGGCTTCGTGCGAATGGGCGCGGT 277
DB      218 CGCGGCTGTGCTTGGGCGCGGCGGAATTCGGGGGCTTCGTGCGAATGGGCGCGGT 277
QY      61 CGTGCGGCTGTGCTTGGGCGCGGCGGAATTCGGGGGCTTCGTGCGAATGGGCGCGGT 120
DB      278 CCGGGTCAAGGGCTGTACCGCTCCCGATGCCCGAGCGGAGCTTATCCGAGACGAGTATG 337
QY      121 CCGGGTCAAGGGCTGTACCGCTCCCGATGCCCGGGGCGGCTTATCCGAGACGAGTATG 180

```

```

QY      338 TTGCAGGCGAGCGGATGACACCTCAGGGACCTTCATGGGACCCCTGCTATGGGGG 397
DB      181 CTGCAGGCTAGCCGAATACACCTCAGGGACCTTCATGGGACCTTCCTGCTATGGGGG 240
QY      398 AACCCCTCAGTCCGACCTGGCTGGCCAGTCAAGGATGATCACTCCCGCAAGACCT 457
DB      241 AACCCCTCAGTCCGACCTGGCTGGCCAGTCAAGGATGATCACTCCCGCAAGACCT 300
QY      458 GCCCTCAGGATTCAGCAGAGTCCAGAGAGGGGTCCAAAATCGAAATCCACAAATGCA 517
DB      301 GCACCTCAACAGATCCAGCAGAGTCCAGAGAGGGGTCCAAAATCGAAATCCACAAATGCA 360
QY      518 AAGAAAAAAGAGATGGCTGACAAAATTTACCTCAAGAGATTCGAACTGGTACAGAA 577
DB      361 AAGAAAAAAGAGATGGCTGACAAAATTTACCTCAAGAGATTCGAACTGGTACAGAA 420
QY      578 TCCAGGCGCTATATGATCTCTTGCTTTTGAAGAAATCGACACAGACTATCATGAGG 637
DB      421 TCAGAGGCGCTACATGATCTCTTGCTTTTGAAGAAATCGACACAGACTATCATGAGG 480
QY      638 AAACGGCTAGATATCCAAAGAGGCTTGAAAGTCCATTAAAGCAAAAACGGAAGCTGCGA 697
DB      481 AAGCGGCTAGATATCCAAAGAGGCTTGAAAGTCCATTAAAGCAAAAACGGAAGCTGCGA 540
QY      698 ATTTTCATTTCAACACTTCAATCCGGCTAAGTCAGATGCGGAGATGGGGAAAGGAG 757
DB      541 ATTTTCATTTCAACACTTCAATCCGGCTAAGTCAGATGCGGAGATGGGGAAAGGAG 600
QY      758 GTGGCTCTCGGAGACTTCCGGTAGAAGAGCGGCTCTGAGAGATTGAGCTTGTCCAAA 817
DB      601 GTGGCTCTCGGAGACTTCCGGTAGAAGAGCGGCTCTGAGAGACTGCTGTCCAAA 660
QY      818 TATGATGCCACTAAACAAAAGAGAGTCTCTTCTTTTAACTCTTGTGATTGAA 877
DB      661 TATGATGCCACTAAACAAAAGAGAGTCTCTTCTTTTAACTCTTGTGATTGAA 720
QY      878 CTGACCAAGAAGCTGTAGGGCCAGACAAACATCTGTAAGATGGCCAGAGCGGCCCT 937
DB      721 CTGACCAAGAAGCTGTAGGGCCAGACAAACATCTGTAAGATGGCCAGAGCGGCCCT 780
QY      938 ACCCAGAGACCGATGGCTTCAGGTGAAGCGGCGCAGAGATGTGAATGTACGCTGTACT 997
DB      781 ACCCAGAGACCGATGGCTTCAGGTGAAGCGGCGCAGAGATGTGAATGTACGCTGTACT 840
QY      998 GTCTACTGATCTGATTAACAGCTTCGCCAGTTTAAATTAGACCCCGCTTACTGGA 1057
DB      841 GTCTACTGATCTGATTAACAGCTTCGCCAGTTTAAATTAGACCCCGCTTACTGGA 900
QY      1058 CTCCTGGGCGATCCATCCCAACCTGCTCACTGATCATCCAAAGCATCTGGCAATATAT 1117
DB      901 CTCCTGGGCGATCCATCCCAACCTGCTCACTGATCATCCAAAGCATCTGGCAATATAT 960
QY      1118 AAGACACATTAAGCTTCAGAGACCTTCAAGCGGGAGTTGTTCATCTGTGACAAAGTACT 1177
DB      961 AAGACACATTAAGCTTCAGAGACCTTCAAGCGGGAGTTGTTCATCTGTGACAAAGTACT 1020
QY      1178 CAGCAGATCTTTAGACTCAACGTAATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCC 1237
DB      1021 CAGCAGATCTTTAGACTCAACGTAATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCC 1080
QY      1238 TTGCTTATGCCACCAAGAACTATCATTCATTCATTCATTCATTCATTCATTCATTCAT 1297
DB      1081 TTGCTTATGCCACCAAGAACTATCATTCATTCATTCATTCATTCATTCATTCATTCAT 1140
QY      1298 CAGAAAAAAGACAGCTTTGATATGACATTAATGTTGAAGTGATGACACTTGAAGACCCAG 1357
DB      1141 CAGAAAAAAGACAGCTTTGATATGACATTAATGTTGAAGTGATGACACTTGAAGACCCAG 1200
QY      1358 ATGAATCTTTTCTGCTGTCCATGCGCAGCCAAAGAGAGATTTGCTTACAGACAA---- 1413
DB      1201 ATGAATCTTTTCTGCTGTCCATGCGCAGCCAGCAGAGAGATTCGCACTTACAGACAAAG 1260

```



```
/db xref="GI:576884"
/translation="PGFPVCSERCRGSLRRSGCCSGPGTGPVPRMGPAPGOG
LYRSPWGAAYRPGMLPGSRMTPOGSPGPNRSGVGRGLAQSGMDORXKRPAP
OOIOYOQOQAVONRNHNNAKKKKWADKILRIRIUREVPEQAAMDLAPRKIDQIMR
KRLDIOELARPIKQKRLRITSTNPNKSDAEDBSTVSWELRGRLEEDAL
SKYDNTQKRRKSSPFKSLVIBUDKOLVDPDNHVEHHTATTQEDGQVRRPGDVN
VACTYLMJLDYOPQPKFLRDLRLGLITQTRPVIIOALMWOIKHKLQDHEBEFV
LCDKYLQOIFESQRMKFSBIPORLHALLMPEPIINHNYSVDPNOCKTACYDIVE
VDDTLKTONSFLSTASQOELATLDNKHETIETINQKTOREFNLSLSPBSGFINDD
WIOSOCRAARR"
polya_signal
3080..3085
/gene="DISK21"
BASE COUNT      657 a      924 c      816 g      704 t
ORIGIN
Query Match      46.4%; Score 1319.4; DB 10; Length 3101;
Best Local Similarity 75.3%; Pred. No. 0;
Matches 2123; Conservative 0; Mismatches 431; Indels 265; Gaps 28;
QY      158 CCGGGCGGGTTCCAGTCTGTGGCTCCAGCGCGCGCGAGCCTCAGAGGGCGGG 217
Db      1 CCGGGCGGGTTTCAATCTGTGCTCCAGCGGTGCCGGGAGCCTCAGAGGAGCGGG 60
QY      218 CCGGGCTGTGCTTGGGCCCAGCGAACTCCGGGCTCTGTGTGCAATGGGCCCGGCT 277
Db      61 CCGGGCGGCTGCTGTGGCCCGGCGGAACTCCGGGCTCTCGGTGCAATGGGCCCGGG 120
QY      278 CCGGGTCAGAGGCTGTACCGCTCCCGATGCCGAGCGGCTATCCAGACCAAGTATG 337
Db      121 CCGGGTCAGAGGCTGTACCGCTCTCCGATGCCGCGGGCGGCTATCCAGACCAAGTATG 180
QY      338 TTGCGACGAGCGCAATGACACTCGAGGACCTTCAATGGGACCCCTGTGGCTATGGGGG 397
Db      181 CTGCGAGGTAGCGCAATGACACTCGAGGACCTTCAATGGGACCTTGTGGGGG 240
QY      398 AACCTTTCAGTCCGACCTGCGCTGCGCCAGTCAGGATGATCAGTCCCGCAAGACCT 457
Db      241 AACCTTTCAGTCCGACCTGCTGTGCGCCAGTCAGGATGAGCAAGTCCCGCAAGACCT 300
QY      458 GCCCTTCAGACGATCCAGACGATCCAGACGAGCGGCTCCAAATGAAACCAATGCA 517
Db      301 GCACCTTCACAGATCCAGACGATCCAGACGAGCGGCTCCAAATGAAATCAATGCA 360
QY      518 AAGAAAAGAAAGATGCTGCTGCAAAATTTCACTCAAGAGATTCGTGACTGTGACAGAA 577
Db      361 AAGAAAAGAAAGATGCTGCTGCAAAATTTCACTCAAGAGATTCGTGAACTGTGACAGAA 420
QY      578 TCCAGGCTATATGATCTCTTGGCTTTGAAAGGAACTGAGACGACTATCATAGG 637
Db      421 TCACAGGCTATATGATCTCTTGGCTTTGAAAGGAACTGAGACGACTATATGAGG 480
QY      638 AAACGGCTATATGATTCAGAGGCTTTGAACGTCCTCATTAAGCAAAAACGAGAGCTGCA 697
Db      481 AAGCGGCTATATATCCAGAGGCTTTGAACGTCCTCATTAAGCAAAAACGAGAGCTGCA 540
QY      698 ATTTTCATTTTAAACCTTTCAATCCGGGTAAGTCAGATGCCAGAGATGGGGAAGGAGG 757
Db      541 ATTTTCATTTTAAACGTTCAATCCGGGTAAGTCAGATGCCAGAGATGGGGAAGGAGG 600
QY      758 GTGGCTTCTGGGAGCTTGGGCTGAGAGAGCGGCTCTGAGAGATTCAGCTTTGCCAA 817
Db      601 GTGGCTTCTGGGAGCTTGGGCTGAGAGAGCGGCTCTGAGAGAGCGGCTTTGCCAA 660
QY      818 TATGATGCACTTAACAAAAGAGAGATTCTCTTCTTTTAACTCTTGTGATGAA 877
Db      661 TATGACGCCACCAAGCAAAAAGAGAGATTCTCTTCTTTTAACTCTTGTGATGAA 720
QY      878 CTGGAACAAAGACCTGTATGGGCAAGCAACATCTGTGATGAGTGGACAGGACCGGCACT 937
Db      721 CTGGAACAAAGACCTGTATGGGCAAGCAACATCTGTGATGAGTGGACAGGACCGGCACT 780
QY      938 ACCCAGGAGACGATGCTTCCAGGTGAGAGCGGCGAGAGATGTGAATAGAGGTACT 997
Db
```

```
Db      781 ACCCAGGAGACCGATGCTTCCAGGTGAGAGCGGCGAGAGATGTGAATAGAGGTACT 840
QY      998 GTCCCTACTGATGCTGATTTACAGGCTCCCGAGTTTAAATTAGACCCCGCTAGCTGCA 1057
Db      841 GTCCCTACTGATGCTGATTTACAGGCTCCCGAGTTTAAATTAGACCCCGCTAGCTGCG 900
QY      1058 CTCTGGGATCATACCGAGCTGCTCAGTATCATCAAGCACTGTGGCAATATAT 1117
Db      901 CTCTGGGATCATATACCGAGCAAGCTCAGTATCATCAAGCACTGTGGCAATATAT 960
QY      1118 AAGACATATAGCTCAAGGACCTCTCAAGAGCGGAGTTTGTATCTGTGCAAGTACTG 1177
Db      961 AAAACACAAGAGCTCAGAGCCCTCAAGAGCGAGAGTTTGTCTGTGACAAATGACTTC 1020
QY      1178 CAGCAGATCTTGAACCTCAAGTATGAGTTTGAAGATTCCTCAGGGGCTCAGTGC 1237
Db      1021 CAGCAGATCTTGAATCTCAAGGATGAGTTTCTAGAGATTCCTCAGGGGCTCAGCGC 1080
QY      1238 TTGCTTATGCGACAGAACTTATCATATTAATCATGTATCATAGTGTGACCCGAATGAT 1297
Db      1081 TTGCTTATGCGACAGAGCCATCATCATCATATCATGTATCATAGTGTGACCCGAATGAT 1140
QY      1298 CAGAAAAGACAGCTTGTATGACATTTGATGTTGAGTGTGACACCTTGAAAGCCAG 1357
Db      1141 CAGAAAAGACAGCGGCTGTATGACATTTGATGTTGAGTGTGACACCTTGAAAGCCAG 1200
QY      1358 ATGAATTTCTTTTCTGTCTCCATCTGCGACCAAGAGAGATTGTCTATACAA---- 1413
Db      1201 ATGAATCTTTCTTTCTGTCTCCATCTGCGACGAGAGAGATTGTCTATACAA-- 1260
QY      1414 ----- 1413
Db      1261 ATTCATGAGACATAGACATCAACAGCTGAAAGACCCAGCAGAGTTCACTGTGAGC 1320
QY      1414 -----CAAGA 1418
Db      1321 TTGCCGAGCCCTCAGTTTCATCAATGATTTGGCTTCAAGTCCAGTGCAGGGCTCAAGA 1380
QY      1419 CAATGATGATGTTGGTTAAACCAAGAGAGAGCGCGAGTGAATTTACTTCCAGC 1478
Db      1381 CGATGATGATGTTGGTTAA-CCGGAAGAGAGCGCTGCTGATGTTCTACTTCCAGC 1439
QY      1479 CTTGGGCTCAGAGGCTGTGTCCGATATCTTTCACTCCAAAGTGTGAGAGACGACAG 1538
Db      1440 CTTGGGCTCAGAGGCTGTGTCCGATATCTTTCACTCCAAAGTGTGAGAGGCGGAG 1499
QY      1539 AATTAGAGCAAGCCCTGGGAATCCGAAATACATAGAGGCTCTCCACAGCCCTGATTGCA 1598
Db      1500 AGTTAGAGCAAGCCCTGGGAATCCGAAACATATAGGCTCT---GTGGCTAGCCTGG 1555
QY      1599 CTGCACCAATTTTGAATTTGGGCTGTGTCTGCTGCTCATATGATCTGCTTGGTCTT 1658
Db      1556 CTGCACCAATTC---CTTGGGCTGTGTCTGCTGCTCATATGATCTGCTTGGTCTT 1611
QY      1659 GCTTGGGCGGTTCCAGGGGATGCTGTGTTCAAGGACAAAGACGAAATGAAAGAGGCT 1718
Db      1612 GCTTGAAGCAATTCAGGGGA---CTTGGCTTCAAGGACAGTGTCAATTAAGAGAGGCT 1667
QY      1719 CACA-----AGACACTGTATTCCTCTTCTTCAACCTTATCTTCCACCC 1766
Db      1668 CACATTTTGTGTCTACAGTCACTGTATCCGCTCTGTATACCCAGTGTGTCCCGCTGCC 1727
QY      1767 CAGCTTCCCTTTGCCCAAAAGTTCCATGTGCTGTACCTCCCTGTGTGTATCATGAG 1826
Db      1728 GTGTGTGTCCCTCCACACCCACCCCGCTCAGCTCCCTCCATAGGCTCTGTGTGTC 1787
QY      1827 ACCTTAGATGTTTGAAGA-GAAGACATGATGAGTAAAGGCTTTGAAATGATG 1885
Db      1788 TCTACCTCCCTATCTTCAATGAGACCTTAGATGATGTTAAGAAACACAGAGTGGGCT 1847
QY      1886 GCCTCAGGACAGTGTGCTTCAAGGGAACAAGCTA-ACTGATCTTACCT--TCAGAGAC 1942
Db      1848 TCTTAGAGTCAAGTGTGTCTTGAAGGAGCAAGCTACATGATCTTGTGCTTGTGAGAGA 1907
```


OY	442	GTCTCCCGAAGAGACTGCTGCCCTCTACAGATCCAGAGGTCCAGACGACGGCGTCCAAA	501
Db	241	GTCTCCCGAAGAGACTGCACTTCAACAATCCAGAGGTCCAGACGACGGCGTCCAAA	300
OY	502	TCGAATCCCAATGCAAGAAAAAGAAATGAGTGGCTGACAAATTTTCACTCAAGAGATTGC	561
Db	301	TCGAATTCACATGCAAGAAAAAGAAATGAGTGGCTGACAAATTTCTACTCAAGAGATTGC	360
OY	562	TGAATGTGTACAGATATCCAGGCTATATGATCTCTTGGCTTTTGAAGAACTTGA	621
Db	361	GGAATGTGTCCAGATACAGGCTCATATGATCTCTTGGCTTTTGAAGAACTTGA	420
OY	622	CCAACATATCATGAGAAACGGCTATGATATCCAGAGGCTTTGAAACGTCCCATTAAGCA	681
Db	421	CCAACATATATGAGAAACGGCTATGATATCCAGAGGCTTTGAAACGTCCCATTAAGCA	480
OY	682	AAAAACGAGGTGGGAATTTCTTCTAACATTTCAATCCGGTATGTACGATCCGCA	741
Db	481	AAAAACGAGGTGGGAATTTCTTCTAACATTTCTTCAATCCGGTATGTACGATCCGCA	540
OY	742	GGATGGGAAAGGACGATGGCTTCCCTGGAGACTTCGGGTAGAAAGACGGCTCTGAGAGA	801
Db	541	GGATGGGAAAGGACGATGGCTTCCCTGGAGACTCCGGGTAGAAAGCCGGCTCTGAGAGA	600
OY	802	TTACAGCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGAAATTCTCTTCTTTAA	861
Db	601	CGCGGCTTGTTCAAATATATGACGCCACAGAAAGAAAGTTCTCTTCTTTTAA	660
OY	862	GTCCTTGGTATTAATGACTGACAAAGACCTGTAATGGGCAAGCAACATCTTGTATGAAATG	921
Db	661	GTCCTTGGTATGACAACTGACAAAGACCTTATAGCCCAACAAACATCTGTATGAAATG	720
OY	922	GCACAGACCGGCACTTACCCAGAGACCGATGGCTTCAGGTGAAACGGCCAGAGAGATGT	981
Db	721	GCAACAGACCGGCACTTACCCAGAGACCGATGGCTTCAGGTGAAACGGCCAGAGAGATGT	780
OY	982	GAATGTACGGGTATCTGTCTACTGATCTGATATACCAGCCTCCCACTTTTAAATTAGA	1041
Db	781	GAATGTACGGGTATCTGTCTACTGATCTGATATACCAGCCTCCCACTTTTAAATTAGA	840
OY	1042	CCCCCGCTAGCTGACTCTTGGGCTATCATACCAGACTGTGCTCCAGTATCATCAAGC	1101
Db	841	CCCCCGCTAGCTGACTCTTGGGCTATCATACCAGACTGTGCTCCAGTATCATCAAGC	900
OY	1102	ACTGTGGCAATATTTAAGACACATAACTCCAGACCCCTACAGACGGGAGTTTGTAT	1161
Db	901	ACTGTGGCAATATTTAATAAACACACAACTCCAGACCCCTACAGACGGGAGTTTGTCT	960
OY	1162	CTGTGACAAAGTACCGGACAGAGATCTTTGAGACTCAACGTAATGAAGTTTTCAGAGATCC	1221
Db	961	CTGTGACAAAGTACCGGACAGAGATCTTTGAAATCTCAGGAGATGAAGTTCTCAGAGATCC	1020
OY	1222	TCAGGAGCTCATGCTTGTCTTATGCAACAGAACATATCATTAATCATGTATCAG	1281
Db	1021	TCAGGAGCTCCACGCTGTTATGCCACAGAGCCCATCATCATCATGTATCATCAG	1080
OY	1282	TGTTGACCCCGATGATTCAGAAAAACAGCTTGTATATACATATGATGTTGAAAGTGATGA	1341
Db	1081	TGTGTACCCCAATGACCAAAAAAGACCGCTGTATATACATTTGACGTGGAGGTGATGA	1140
OY	1342	CACCTTGAAAGCCCGATGAATTTCTTTCTGTGTCACATCCAGCCCAACAGAGAGATTGC	1401
Db	1141	CACCTTGAAAGCCCGATGAATCTTTTCTGTGTCCACTCCAGCCCAACAGAGAGATGC	1200
OY	1402	TACTCTAGACA-----	1413
Db	1201	CACCTTACACAACAAGATTCATAGAGCATAGACATCAACACGTTGAAGACCCAGCG	1260
OY	1414	-----	1413
Db	1261	AGAGTTCATGTTGAGCTTTGCCCGAGACCTCAGGGTTTCATCATATGATGGCTTCAATGC	1320

QY	1414	-----CAAGCAATGACGATGTGTGGGTAAACC	CAAGGAGGAGCGCG	1458
Db	1321	CCAGTGCAGGAGCCTCAAGAGACTGATGTGTGGTAA	CCCGAAGGAGCGCTCG	1380
QY	1459	AGCTGAGTTCTACTCCAGCGCTTGAGGCTCAGAGAGCTGTGTGGCATACTTCTACTCAA	1518	
Db	1381	TGCTGAGTTCTACTCCAGCGCTTGAGGAGCTGTGTGGCATACTTCTACTCAA	1440	
QY	1519	GGTGCAGCAGAGCAGCAGAGATTAGACCAAGCC	CTGGGAATCCGAAATCAATAGGCGCT	1578
Db	1441	GGTGCAGCAGAGGCGGCAAGATTAGAGCAAGCC	CTGGGAATCCGAAACATATAGGCGCT	1500
QY	1579	CTCCCAAGCGCCCTATTTGACTGCACCAATTTCTGATT	TGGGCGCTGTGCTGCTGCTC	1638
Db	1501	CT---GTGGCCCTTAGCGCTGCACCGATTC---CTTGGGCGCTGTGCTGCTGCTC	1553	
QY	1639	ATTAGTATCTGCTTGTGCTTGTGCTTGGGAGCGTTCCAGGAGATGCTGTTGGTTCAAGGACAA	1698	
Db	1554	AGTGTACTGTCTTGTGCTTGTGCTTGAAGGCAATTCAGGAGGA---CTTGGCTTCAGGACAG	1609	
QY	1699	GACCAGAAATGAAGAGGTCTCACA-----AGACACCTGTATCCTCTTCTTTC	1746	
Db	1610	TGTGCATATGAAGAGGGGTGTCACATTTCTGTCTCAGACTACCTGTTATCCGTCCTGTGA	1669	
QY	1747	ACCCATATCTTCTTCCACCCCGACTTCCCTTTGGCCCCA	CAAAGTTCCCATGTGCTGTAC	1806
Db	1670	CCCCAGTCTGCCCCCGTCCGTCGTGTCCCCCTCCACCCACCCCGCTCAGCTCTCTCC	1729	
QY	1807	CCCTCCCTCGTCTACATAGGACCTCTAGATATGTATGTATGAGAGAGAAC--ATGTAGTGTGA	1864	
Db	1730	CCATGAGGCTCTGTGTGCTCTTACCTCCATCTACATATGAGACCTCTGATATGTGTTA	1789	
QY	1865	ATGATGTCTTGGAATGATTGGCCCTCAGGCGAGTGTGCTTTC	CAAGGAGACAGCTA-ACT	1923
Db	1790	GAGAACACAGAGTGGGGGCGCTCTGAGTGCAGGTGTCTTGAAGGAGACAGACTACACT	1849	
QY	1924	GATCCTTACCT--TCAAGAACCCAGAGATTGGTTTGCTTCTTCC	AAAGTCTCAGGCC	1961
Db	1850	GATCCTGCTCCCTTGTCCAGAGACCTTAGGCGCTTGGAGCATATCCCTGTCTGAGCCTCAAGGCC	1909	
QY	1982	TGTGGGCACTTATAAGCTATGTGATCTTGGCTCTCTGTATATCAGAAATCCAATTTCTT	2041	
Db	1910	T-AGGGGAGTGTGAACCTAGCTAGCACTTGGCCCTCCCGTAGCTTGAC---TTCTTC	1963	
QY	2042	CCTTCCCTCCACAGAGTTTGGAACTCAACCTCCCTTCACTTGTGTCCTGTAAGCACTACAG	2101	
Db	1964	CCTCCCTCCGAGAGTTGGGGCAGAGG-----CTCCTTACCTCTGAGAGTAAAG	2013	
QY	2102	AAACCTGTGTTCTTGGCTCACTAGAGCCGAGGTCAGTCCCAAGCCCTTGGGTTGAGCTG	2161	
Db	2014	GAGC-----CTGGGCTTACATGAGCCCGGGGTGTCCCTCTCCCTCTGGAATTAAAC	2066	
QY	2162	CTGTCAGTGCTTCTCTCACTCTTAAATTGGGGTTCACATACATATTGGA--GTTTGTTC	2219	
Db	2067	TGCTGTCTCAAGTGTCTCTGACCCCTTAGGGGTTCAGATTAATGAGATGTGTGTGA	2126	
QY	2220	TTTATTTGCTCCCTCCACAGACACTCCCTGTGGGCTGCCCTTGTGATTCCTTCAGATCTGCC	2279	
Db	2127	ATTGTGTGCTCTCTCCACACACTCTCCGTAGCCGCCCATTTAGGATTTCCCTTACACTGC	2186	
QY	2280	CTAATCCCGGCAATTTGGGTGGGGAATCTTGTGCTTTCCCTTTCAGAGCCCAAGGATCT	2339	
Db	2187	CTTAACCAACGCTTTTGGTTGGGG-ATCTTGGCTTTCTT-----	2225	
QY	2340	CATCTGGGAACTGTCAATTGCGCAGCAGAGGCTGTTCTTCTCTGCAGTTTGGAGATGTGAC	2399	
Db	2226	-----TGTCATTTCCACAGCAAGACTGTTCTTCCGCTGTATTAAGGAGTGTGCT	2273	
QY	2400	TCATTTCATTCACATCACTACACCTGTGCTGTGACATCCCTTAATGAGAGAAACGGGCTTAA	2459	
Db	2274	TGTTT-----ATTCACTTCCACCTGTGCCCCCTCTCG--TAAATGGAAGAAACAGGCTGAA	2325	
QY	2460	ACCAAAAGGATAAAAAGCCCTGGGCAATCCCTGTCTTCTGTGCCCTTGTCTGTGCCAGTT	2519	

```

Db      2326  ATCAAAACGGGT---AAAGCCCTAGGCGCATCCCTGTCTTCTGTCCTCATGTCTGCCAGTT 2382
Qy      2520  GACACCTACTGTGACTTCTAGGCGACATCAGGAGTGAAGCCCTAGGCGTGGAGAAATAG 2579
Db      2383  GAATCCCACTGTGTGCTTCCCGGCGACTGAGGAGTAAAGCCCTAGGCGTGGAGAAATAG 2442
Qy      2580  CGCTCAGTTGGTTGTGACTCTTCCCTCCTCCCTGCTCACAGGATTGTGACTCCCCAGC 2639
Db      2443  GTCTGAATGGTTGTGACTC-----CCACCCCCCTG 2475
Qy      2640  CCTCGCCTCAAAAGTTTACAGCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGCCCT 2699
Db      2476  CCTCGCCTCAAAAGTTTACAGCCCTCAGGAGCAGCAGGA-TGTGGATCGAGGCCCT 2534
Qy      2700  TGGATCTGAGATGTTTGTGATCTTCCAGAGAGCCTCACATCTTCTTCCAGTTGT 2759
Db      2535  TGGGACAGA-----TGCCTTGAATCTTCCAGGGAAGCCTCCGATCTTTC-----CAGTT 2584
Qy      2760  ATCAACCCCGAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGTTGGAGACA 2819
Db      2585  TGTACCCCGAGTTAGCATGTCCAGGCTCGCAGACACACAC---TGCAGGTTGGAGACA 2641
Qy      2820  GCTGGGCACAAAGG 2834
Db      2642  GCTGGGCACAGGGG 2656

RESULT 8
AC025154
LOCUS
DEFINITION Homo sapiens 12 BAC RP11-469H8 (Roswell Park Cancer Institute Human
AC025154
VERSION
KEYWORDS HTG.
SOURCE AC025154.31 GI:22549680
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164813)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M.P., Bryant,N.P.,
Buahy,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
Escotto,C., Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P.,
Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M.,
Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Toshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Koryah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
Marcondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

```

```

Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shim,C., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 164813)
Worley,K.C.
Direct Submission
Submitted (07-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 164813)
Worley,K.C.
Direct Submission
Submitted (26-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 164813)
Worley,K.C.
Direct Submission
Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 164813)
Worley,K.C.
Direct Submission
Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 164813)
Worley,K.C.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
7 (bases 1 to 164813)
Worley,K.C.
Direct Submission
Submitted (15-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 30, 2002 this sequence version replaced gi:22213007.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STGs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

```

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum

Db 489 GACCATCATGCGCAACGACTGTGATATTAGGAAGCTCTCAAGAGACCCATTAAGCAAAA 548
Qy 685 ACGAAGCTGCGAATTTTTCATTCTTAAACATTTTCAATCGGCTAGTACAGATGCGGAGA 744
Db 549 AAGAAAATCCGAATATTATCTCCACACATTTCAACCTCGCAAGCCAGATGACAGAGA 608
Qy 745 TGGGGAAGGAGCGGTGGCTTCTGGAGCTTGGGTAGAAAGACGGCTCTCTGGAGGATTC 804
Db 609 TGGCGAGGAACTGTTCATCTTGGAGTTCGCTGTGGAGGCGGCTCTCTGGAGGATTC 668
Qy 805 AGCCTTGTCCAAATATGATGCACTAAACAAAAGAGGAGTCTCTCTCTCTTTTAAGTC 864
Db 669 TGCTGTGTCCAGTATGAAGCACCACCAAGCAGAAGAGGAGTCTCTCTCTTTTCAAGTC 728
Qy 865 CTGTGTGATTGAACCTGGAACAAGACCTGTATGGCCAGACAACATCTGTAGATGCA 924
Db 729 TTTAGTGATGAGTTGGACAAAGACTTGTATGGACCTGACATCATTTTGGTTGAGTGCA 788
Qy 925 CAGGACCCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGCCAGGAGATGTGAA 984
Db 789 CAGAACACGCCAACACTCAGGACACAGATGGATTTTCAGGTCAAGAGCCCTGGGATGTTG 848
Qy 985 TGACGCTGTACTGTCTACTGATGCTGATTAACAGCTCCAGCTCCAGTTTAAATAGACC 1044
Db 849 GGTCCGCTGTACTGTACTGTCTCATGCTGACTACCAAGCCCTCAGTATAGCTGGACCC 908
Qy 1045 CGCCTAGCTGACTCCTCGGCGATCCATACCCAGACTCGTCCAGTGAAGCGCCAGGAGACT 1104
Db 909 CGCCTGCGCGCTGCTGGGATTTCACTCAGACCCCGCTGTGATCATTCAGGCTTT 968
Qy 1105 GTGGCAATATATAAGACACATAAGCTCCAGGACCTCAGAGCGGGAGTTTGTCTATCTG 1164
Db 969 GTGGCAGTATGTGAAGACCCACAACTCCAGGACCTCATGAACAGAGTTTCAATCACTG 1028
Qy 1165 TGACAAATACCTCCAGCAGATTTTGAACCCAGAGGATGAAGTTCTCTGAGATCCCTCA 1088
Qy 1225 CGGCTCCATGCTGCTTATGCCACAGAACCTATCATTAATCATGTCATCAGTGT 1284
Db 1089 CGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1148
Qy 1285 TGACCCGAATGATCAAAAAAGACAGCTTGTGTATGACATGATGTTGAAGTGGATGACAC 1344
Db 1149 TGATCCCAATGATCAAGAGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1208
Qy 1345 CTTGAAGACCCAGATGAATCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1404
Db 1209 CTTGAAGACCCAGATGAATCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1268
Qy 1405 TCTAGACACAGA 1418
Db 1269 CTTGACACACAGA 1282

RESULT 11
BT007694
LOCUS BT007694.1 1413 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3 mRNA, partial cds.
ACCESSION BT007694
VERSION BT007694.1 GI:30584226
KEYWORDS FLI_CDNA.
SOURCE synthetic construct
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1413)
AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M., and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor vector

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1413)
AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M., and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Last distribution: http://bioinfo.clontech.com/orfclones.
Clone Location/Qualifiers
1..1413
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH000911.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/note="vector: pDNR-Dual"
1..>1413
/note="Mutations: 12:OK;1412:Stop->Leu"
/codon_start=1
/transl_table=1
/product="Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3"
/protein_id="AAP36362.1"
/db_xref="GI:30584227"
/translation="MTGLQHPPTVORPMPGSGARPHOGAMPGPSPYMGSPAVR PLGAPAGMEAPKRAAPPPGSOAQSGOPVPTAPARSRAKRRKMDKIUPORIEL VPESQAYMDLLAFERKLDITMRKVDIOEALKEPMKOKRLVYISNTFNPAKPD AE DSQSIASWELRVEGKLLDPPSKOKRFSSFFKSLVIELDKDLYGPNHNLVEMRTPT TQETDGFQVKKRPGDLVSRCTLLMLDYPFPKLDPLRLGLLHSTRSRAIVQALWQ YVKNRLQDSDHKEYINGDKYFQIFDCPRLKFSIEIPORTALLPLPDPPIVNHVIS DPSQDKTACVDIDVEVEPLKQMGSLSTANQOBSALDSKIHETIESINOLKIQ RDMFLSFRDPKGVVODLLRSQSDRLKVMTDVAGNPBEERRAEFVHQPSQEAVSRYF YCKIQQRQBLEQSLVVRNTL"
BASE COUNT 320 a 446 c 404 g 243 t
ORIGIN
Query Match 18.9%; Score 537.6; DB 12; Length 1413;
Best Local Similarity 70.2%; Pred. No. 1.3e-127;
Matches 786; Conservative 0; Mismatches 289; Indels 45; Gaps 3;
Qy 323 CCGAGACAGGATGTTGCCAGGACGCGAATGACCTCAGGACCTTCATGGGACCC 382
Db 37 CAGCGCCCGGGATGCGCTCTGGAGCCCGGATGCCCAACAGGGGGGCCCATGGGCC 96
Qy 383 CTTGGC-----TATGGGGGAACCTTCAGTCGACCTGGCCCTGGCCAGTCAGGATG 436
Db 97 CCGGGCTCCCCGTACATGGGACGCGCCCGCTGCGACCCCGCCCTGGCCCGGGCATG 156
Qy 437 GATCAGTCCCGCAAGAGACCTGCCCTTCAG-----CAGATCCAGCAG 478
Db 157 GAGCCCGCCCGCAGGAGCAGCGCCCGCCCGGGGAGAGCCAGGACAGCCAGGCGC 216
Qy 479 GTCCAGCAGCAGGCGGTCCAAAATCGAAACCAATGCAAAAGAAAAGAGAGTGGCTGAC 538
Db 217 CAGCGGTGCGCACCGCCCGCGCGGAGCGCGAGTGCACAGAGAGAGAGAGTGGCTGAC 276
Qy 539 AAAATTCTACCTCAAGGATTCGTGAAGTGTGTACAGATCCAGGCTTATGATGATC 598
Db 277 AAAATCTCTCTCAAGGATTCGGAGCTGTGTCCTCCGAGTGTGTCCTCCAGGCTTATGATGATC 336

599 TTGGCTTTTGAAGAAAGTGAACAGATCATATGAGAAAAGCGTAGATATTCAGAG 658
 337 TTGGCTTTTGAAGAAAGTGAACAGATCATATGAGAAAAGCGTAGATATTCAGAG 396
 659 GCCTTGAAGAAAGTGAACAGATCATATGAGAAAAGCGTAGATATTCAGAG 718
 397 GCCTTGAAGAAAGTGAACAGATCATATGAGAAAAGCGTAGATATTCAGAG 456
 719 AATCCGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 778
 457 AATCCGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 516
 779 GTAGAAGAGCGAGTCTCTGAGAGTTCAGCTTGTCCAAATATGATGCCACTAAGAAAAG 838
 517 GTAGAAGAGCGAGTCTCTGAGAGTTCAGCTTGTCCAAATATGATGCCACTAAGAAAAG 555
 839 AGGAAGTCT 898
 556 CGGAAGTCT 615
 899 CGAGAACCAATCTGTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 958
 616 CTGACCAACCAATCTGTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 675
 959 CAGTGAAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1018
 676 CAGTGAAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 735
 1019 CAGCT 1078
 736 CAGCT 795
 1079 ACTGCTCCAGTATATATCAACCAATCTGTGATGAGTGAAGTGAAGTGAAGTGAAG 1138
 796 AATGCTCCAGTATATATCAACCAATCTGTGATGAGTGAAGTGAAGTGAAGTGAAG 855
 1139 CTTCAAGAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1198
 856 TCCCATGACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 915
 1199 CTTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1258
 916 CCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 975
 1259 ATCATCAT 1318
 976 ATGTCATCAT 1035
 1319 GATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1378
 1036 GATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
 1379 ACTGCGAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1418
 1096 ACGGCCAACCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1135

RESULT 12
 BC002628 1698 bp mRNA linear PRI 12-JUL-2001
 LOCUS Homo sapiens, similar to SWI/SNF related, matrix associated, actin
 dependent regulator of chromatin, subfamily d, member 3, clone
 MGC:3653 IMAGE:3608512, mRNA, complete cds.
 ACCESSION BC002628.1 GI:12803586
 VERSION BC002628
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1698)
 AUTHORS Strausberg, R.
 TITLE Direct Submission

JOURNAL
 Submitted (05-FEB-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nigri.nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McElroy, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stancir, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H., and Green, E.D.

FEATURES
 source
 1..1698
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:3653 IMAGE:3608512"
 /tissue_type="uterus, endometrium adenocarcinoma"
 /clone_id="NIH MGC 44"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 139..1551
 /codon_start=1
 /product="similar to SWI/SNF related, matrix associated,
 actin dependent regulator of chromatin, subfamily d,
 member 3"
 /protein_id="AA02628.1"
 /db_xref="GI:12803587"
 /translation="MTPLQHPPTVVOPEPSPGARNPHGAPMGPPSPYWGSPAVR
 PGLAPAGMEPRKRAAPPGQSOQSPPTAPARSRARRRMADKILPRLRL
 VPSQAVMDLAFRRKLDOTIMRKRVDIOELKRPKOKRLRLVTSNTPADKDAE
 DSDGSIASWELRVGKLLDPSKOKRSPFSKSLVLELDLGLPDNHLVEMHPT
 TOERDGPVOKRPRDLRSVCTLLMLNDPOPPKLPRLRLGLHTGSRSAIYQALWQ
 YKTNRLDSDHDKETINDKRYFOQIFDCKRSEIPLRLALPLPPPIYINHTISV
 DPSQKNTACVDIVVEEPLKQMSFLSTANQOETSLDLSKIHETESINDKIQ
 RDPMLSFSDRDKGVODLRLRSQSRDLKVMYDVANPEERAEFEYHQMSQEAVERFV
 YCKIQRROELQEOSLVAVNT"

CDS
 139..1551
 /codon_start=1
 /product="similar to SWI/SNF related, matrix associated,
 actin dependent regulator of chromatin, subfamily d,
 member 3"
 /protein_id="AA02628.1"
 /db_xref="GI:12803587"
 /translation="MTPLQHPPTVVOPEPSPGARNPHGAPMGPPSPYWGSPAVR
 PGLAPAGMEPRKRAAPPGQSOQSPPTAPARSRARRRMADKILPRLRL
 VPSQAVMDLAFRRKLDOTIMRKRVDIOELKRPKOKRLRLVTSNTPADKDAE
 DSDGSIASWELRVGKLLDPSKOKRSPFSKSLVLELDLGLPDNHLVEMHPT
 TOERDGPVOKRPRDLRSVCTLLMLNDPOPPKLPRLRLGLHTGSRSAIYQALWQ
 YKTNRLDSDHDKETINDKRYFOQIFDCKRSEIPLRLALPLPPPIYINHTISV
 DPSQKNTACVDIVVEEPLKQMSFLSTANQOETSLDLSKIHETESINDKIQ
 RDPMLSFSDRDKGVODLRLRSQSRDLKVMYDVANPEERAEFEYHQMSQEAVERFV
 YCKIQRROELQEOSLVAVNT"

BASE COUNT 386 a 535 c 480 g 297 t
 ORIGIN

Query Match 18.9% Score 537.6; DB 9; Length 1698;
 Best Local Similarity 70.2%; Pred. No. 1.4e-127;
 Matches 786; Conservative 0; Mismatches 289; Indels 45; Gaps 3;
 323 CCGAGACCAAGTATTTCCAGGAGCGGAGTGAACCTTCCATGGAGACCC 382
 175 CAGCGCCCGGGAGTCCCTGTGAGCCCGGATGCCCCACAGGGGGCCCATAGGGCC 234
 383 CTTGGC-----TATGGGGGAACCTTCAATCGACTGCGCTGGCCAGTCAAGGATG 436
 235 CCGGCTCTCCGATCATGGAGAGCCCGCGCTGGACCCCGGCTGGCCCGCGGGCATG 294
 437 GATCAGTCCGCAAGAGACTGCGCTTCAG-----CAATCCAGAG 478
 295 GAGCGCCCGCCAGAGCGAGCGCGCCCGCGGGGAGAGCCAGGACAGAGCGAGGCG 354

QY 479 GTCCAGCAGCGCGGTCCAAAATCCAAATGCAAAAGAAAAAGAGATGCTGAC 538
 Db |||||
 355 CAGCGGTGCCCCACCGCCCCCGCGGAGCGCAGTGCACAGAGAGAGATGCTGAC 414
 QY 539 AAAATTCCTACCTCAAGGATTCGTGAACCTGGTACCAGATCCAGGCTATATGATCTC 598
 Db |||||
 415 AAAATCCTCCCTCAAGGATTCGGAGCTGGTCCCGAGTCCCGAGCTTACATGACCTC 474
 QY 599 TTGGCTTTTGAAGGAATCTGACACAGACTATCATGAGGAACGGCTAGATATCCAAGAG 658
 Db |||||
 475 TTGGCATTTGAGAGGAATCTGATCAACCATCATGCGAAGCGGTGGACATCCAGGAG 534
 QY 659 GCCTTGAACGCTCCCATTAAGCAAAAACGGAAGCTCGCAATTTTCAATCTTCAACACTTTC 718
 Db |||||
 535 GCTCTGAAGAGGCCCATGAGCAAAAGCGGAGCTCGCACTATATCTCCAACACTTTT 594
 QY 719 AATCCGGCTAAGTCAAGTCCGAGATGGGGAAGGAGCGTGGCTTCTGGAGCTTGG 778
 Db |||||
 595 AACCCCTGCGAAGCCTGATGCTGAGGATTCGCGAGCAGCATTCGCTCTGGAGCTACGG 654
 QY 779 GTAGAAGGACGGCTCCTGGAGGATTCAGCCTTTGTCCAAATATGATGCCACTAAACAAAAG 838
 Db |||||
 655 GTGGAGGGAAGCTCCTG-----ATGATCCCGAGCAACAGAG 693
 QY 839 AGGAAGTTCTCTCTCTCTTTAACTCTTTGATTTGAACCTGGACAAAGACCTGTATGG 898
 Db |||||
 694 CGGAAGTTCTCTCTCTCTTCAAGAGTTTGGTATCGAGCTGGACAAAGATCTTTATGGC 753
 QY 899 CCAGACAACCATCTGTTAGATGGACAGGACCGGCACTACCAGGAGACCGATGGCTTC 958
 Db |||||
 754 CCTGACAACCATCTGTTGAGTGGCATCGGACACCCAGCACCGAGGACGCGGCTTC 813
 QY 959 CAGGTGAAGCGCCAGGAGATGTAATGACGGTGTACTGCTCTACTGATGCTGATTAC 1018
 Db |||||
 814 CAGGTGAAGCGCCCTGGGACCTGAGTGGCCCTGCAGCTCTCTCATGCTGGACTAC 873
 QY 1019 CAGCCTCCCCAGTTTAAATTAGACCCCCCGCTAGCTCGACTCCTGGGCATCCATCCAG 1078
 Db |||||
 874 CAGCCTCCCCAGTTTCAAACCTGGATCCCCGCTAGCCGCTGCTGGGCTGCACACAG 933
 QY 1079 ACTGTCAGTGATCATCAAGCACTGGGCAATATTAAGACACATAAGTCCAGGAC 1138
 Db |||||
 934 AGCCGCTCAGCCATTTGTCAGGCCCCGTGGCAATATGTGAAGACCAACAGGCTGCAGGAC 993
 QY 1139 CCTCAGGAGCGGAGTTTGTATCTGTGACAGTACCTGCAGCAGATCTTTGAGACTCAA 1198
 Db |||||
 994 TCCCATGACAAGGAATACATCAATGGGACAGATATTTCCAGCAGATTTTGTATGTCCTC 1053
 QY 1199 CGTATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTTCCTTATGCCACCAAGCT 1258
 Db |||||
 1054 CGGCTGAAGTTTCTGAGATTTCCAGCGCCTCAGCGCTCTATTTGCCCTGACCCA 1113
 QY 1259 ATCATCATTAATCATGTCATAGTGTGACCGAATGATCAGAAAAGACAGCTTGTAT 1318
 Db |||||
 1114 ATTGTATCAACCATGTATCAGCGTGACCCCTTCAGACCAGAAAGACGCGTGTAT 1173
 QY 1319 GACATTGATTTGAAGTGTGATCAGCTTGAAGCCAGATCAATCTTTCTGCTGTC 1378
 Db |||||
 1174 GACATTGAGTGGAGTGGAGAGCCATTAAAGGGGCGAGATGAGGAGCTTCTCTCTATCC 1233
 QY 1379 ACTGCCAGCAACAGGAGATTCCTACTCTAGACAACAAGA 1418
 Db |||||
 1234 AGGCCCAACACAGAGATCAGTCTCTGGACAGTAAGA 1273
 Db |||||
 RESULT 13
 BC013122
 LOCUS
 DEFINITION
 Mus musculus SWI/SNF related, matrix associated, actin dependent
 regulator of chromatin, subfamily d, member 3, mRNA (cdna clone
 MGC:8018 IMAGE:3586276), complete cds.
 ACCESSION
 BC013122
 VERSION
 BC013122.1 GI:15341882

KEYWORDS SOURCE ORGANISM

REFERENCE AUTHORS

MGC.
 Mus musculus (house mouse)
 Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1790)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.D.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schuruz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 1790)
 Strausberg, R.
 Direct Submission
 Submitted (27-AUG-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisegeed, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 10 Row: 1 Column: 23
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 13385379.

FEATURES source

1..1790
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:8018 IMAGE:3586276"
 /tissue_type="Mammary tumor. Metallothionien-TGF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam1"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..1790
 /gene="Smardc3"
 /db_xref="LocusID:66993"
 /db_xref="MGI:1914243"
 488..1633

gene

CDS

```

/codon_start=1
/product="Smardc3 protein"
/protein_id="AAH13122.1"
/db_xref="GI:15341883"
/db_xref="LOCUSID:66993"
/translation="MADKILPQIRIELVPEQAVMDLLAFERKLDOTIMRKRVDIQEA
LKRPMKQKRLKLYISNTENPAKPDABSDGSIASMEKREVEKLDLDDSKQKRFSSP
FKSLVIELDKDLYGPDNHLVEMHRTPTQETGDPQKREKLDLSDVETLLMDYQPO
FKLDPRLARLGLHDSQSAIYQALMOYKRNRLQDSHDKETIDVNDKXEQOITFCBPRL
KRESEIPQRLTALLPDPPIVINHVTISVPSDQKTRACDIDIVEEPEPLKQMSFLLS
TANQELIALDSKLIHETLESINQKIQIDPFMSFSRDPGVQDILRSQKRLKVMTD
VAGNEERERRRPFYHQPWSEAVSRYPFYCKIQORQEBLEOSLVENT"
BASE COUNT      447 a      522 c      516 g      305 t
ORIGIN
Query Match      18.7%; Score 530.8; DB 10; Length 1790;
Best Local Similarity 69.9%; Pred. No. 8e-126;
Matches 781; Conservative 0; Mismatches 292; Indels 45; Gaps 3;
QY      325 GAGACCGAGTATGTCGCGAGCGAGATGACCTTCAGGGACCTTCATGGGACCCCC 384
DB      259 GGGCCCCGGGATGCGCTGAGACCCGAATGCCCAACGAGGGGGGCCCATGGCCCC 318
QY      385 TGGC-----TATGGGGGGAACCTTTCAGTCCGACCTGGCCCTGAGTCAGGATGGA 438
DB      319 GGGGTCGCCCGTACATGGGACGCCCGCGGTACGACCCGGCCCTGGCCCCGGGATGGA 378
QY      439 TCAGTCCGCAAGAGACTGCTCCCTCAG-----CAGATCCAGAGGT 480
DB      379 GCCCGCCGCAAGCAGCAGCGCCCGCCCGGCGAGAGCCAGGACAGGGCCAGGCGCA 438
QY      481 CCAGACGACGAGGGGTCCTCAAAATCGAAACCAATGCAAGAAAGAAAGATGGCTGACAA 540
DB      439 GCCCGTCGCCACCGCCCGGAGCGGAGTCGACGAGAGAGAGAAATGGCTGACAA 498
QY      541 AATTCTACCTCAAGAGATTCGTGAATGCTGACAGATCCGAGCCCTATATGATCTCTT 600
DB      499 AATCTCTCTCAAGAGATTCGGAGCTGATACCCAGTCCGAGCTTAATGACCTCTCT 558
QY      601 GCGCTTTGAAGAGAACTGAGACCAATATCATGAGAAACGCTATGATATCCAGAGGC 660
DB      559 AGCATTTGAGAGAGAACTGATCAACATCATGCGAGCGGTGATCATCGAGAGGC 618
QY      661 CTTGAAAGCTCCCATTAACCAAAAGAGCTGGCAATTTTCATTCTTAACATTTCA 720
DB      619 CTTGAAGAGCCCATGAAAGCAAGGAAAGCTGCGCTTATATCTCAAAATACCTTTAA 678
QY      721 TCCGGCTAGTCAATGCGGAGATGGGAGAGGACGATGCGCTTCTGGAGCTTCGGGT 780
DB      679 CCTCGGAAGCTGATGCGGAGACCTCTGATGACACATTCCTCTGGAGAGCTCGGGT 738
QY      781 AGAAGAGCGCTCTCGAGAGATTCAGCTTGTCCAAATATGATGCCATTAACAAAGAG 840
DB      739 GGAGGGGAGACTCTTGG-----ATGATCTAGTAGAGAGAG 777
QY      841 GAAATCTCTTCTCTTTAAGTCTTGTGATTAACCTGAGCAAGACCTGATGAGGC 900
DB      778 GAAGTTTCTCTCTCTTCAAGAGTTGTGATGAGTTGAGCAAGACCTTATATGCGCC 837
QY      901 AGACAACCATCTGTAGATGCGACAGACCGGCACATACCCAGAGAGACGATGGCTTCA 960
DB      838 AGACAACCATCTGTGTAGTGGACCGACACCCAGAACCCAGAGAGAGAGGTTTCA 897
QY      961 GGTGAAGCGGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB      898 AGTGAAGAGACCGAGGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 957
QY      1021 GCGTCCCGAGTTAATTTAGACCCCGCTAGTGTGATCTCTGGGATCATACCAAGAC 1080
DB      958 GCGTCCCGAGTTAATTTAGACCCCGCTAGTGTGATCTCTGGGATCATACCAAGAG 1017
QY      1081 TGGTCAGGATCATCGAGACCTGAGCAATATATTAAGACACATAGCTCAGAGACCC 1140

```

```

DB      1018 CCGCTACCGCATTTGTCACGAGCACTGTGCGAGTATGTGAAGCAACAGGCTACAGACTC 1077
QY      1141 TCAGAGGGGGAGTTTGTTCATCTGTGACAGAGTCTGAGAGATCTTGTGACTCAAGC 1200
DB      1078 CATGACAGAGAGTATCATTAATGCGACAGATATTTCAGAGATTTTGTACTGCCCGC 1137
QY      1201 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTGTATGCCAGAGAACTAT 1260
DB      1138 CCTAAGTTCTGTAGATTTCCCGAGCGCTCAGAGCCCTGCTGCTGCCCGTCAACAT 1197
QY      1261 CATCATTAATATGATCATGAGTGTGACCCGAGATGATGAGAAAAGACAGCTTTATGA 1320
DB      1198 TGTGATCAACACCGTCATCAGCTGTGAGCCCATCAAGACAGAGAGACAGCGCTATGA 1257
QY      1321 CATGATGTTGAGTGTGACATCCTTGAAGCCAGATGATTTCTTTCTGCTGTCAAC 1380
DB      1258 CATGATGTGAGGAGGAGGAGAACCGCTAAGAGAGATGATGATCTCTCTGTCCAC 1317
QY      1381 TGCAGCCACAGAGATTTGCTACTCTAGACAAACAGA 1418
DB      1318 GGCACACAGACAGAGATCAGTCTCTGTGACAGTAA 1355

RESULT 14
HSU66619      1724 bp      mRNA      linear      PRI 18-SEP-1996
LOCUS        Human SWI/SNF complex 60. Kda subunit (BAF60c) mRNA, complete cde..
DEFINITION   U66619.1 GI:1549246
ACCESSION    U66619.1
VERSION      1
KEYWORDS     Homo sapiens
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.,
              (bases 1 to 1724)
              Diversity and specialization of mammalian SWI/SNF complexes
              Genes Dev. 10 (17), 2117-2130 (1996)
              PUBMED 8804307
              MEDLINE 8804307
TITLE        Purification and biochemical heterogeneity of the mammalian SWI-SNF
              complex
              EMBO J. 15 (1996) In press
              2 (bases 1 to 1724)
AUTHORS      Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.
JOURNAL      Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE      PUBMED 8804307
PUBMED       8804307
REFERENCE    Wang, W., Xue, Y., Zhou, S. and Crabtree, G. R.
              Direct Submission
              Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford
              University, Beckman Center B207, Stanford, CA 94305-5428, USA
FEATURES     Location/Qualifiers
              1..1724
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /cell_type="Jurkat T cells"
              1..1724
              /gene="BAF60c"
              181..1590
              /gene="BAF60c"
              /note="Unlike BAF60a and BAF60b this gene is highly
              expressed in muscle cells; similar to human BAF60a,
              BAF60b, BAF60c and yeast Swf73 and Swf73b"
              /codon_start=1
              /product="SWI/SNF complex 60 Kda subunit"
              /protein_id="AAC50697.1"
              /db_xref="GI:1549247"
              /translation="MTPEGLHPPTVVGPRGPMGARMHGRPMAPRTWAAPCD
              PAMPRPDARGAASAPAROSQAOQGPPTAPARSRSAKRKKADILPQIRIELV
              PESQAVMDLLAFERKLDOTIMRKGVDIQELAKRPMKQKRLKLYISNTENPAKSD
              SDGSIASMEKREVEKLDLDDSKQKRFSSPKSLVIELDKDLYGPDNHLVEMHRTPTT

```

QETDGFQVKPGLSVRCCTLLMLDVPQPKLPDRLPRLGLLQTSRSIAVALMQY
VKNRLQDSHKEYINGDKYFOQIEDCPRLKFSI PQRLTALLPPDPIVNHVISVD
PSQKKTACVDI DEVEEPLKQMSGSLSTANQOETSPDLSKTHEITESINGKJQR
DFWLSRDRPKGVQDLLRSQSRDLKWTVDVANGPEERAEFFHQWPSQEAWSRYFY
CKTCOORRQEGVSLVVRNT"

BASE COUNT	376 a.	546 c	510 g	292 t
ORIGIN				

Query Match 18.4%; Score 522.8; DB 9; Length 1724;
Best Local Similarity 69.9%; Pred. No. 9.4e-124;
Matches 782; Conservative 0; Mismatches 292; Indels 44;

Qy	323	CCGAGACCAAGTATGTTGCCAGGCAGCCGGAATGACACCTCAGGACACCTTCATATGGGACCC	382
Db	217	CAGCGCCCCGGGATGCGGTCTGGAGCCCGAATGCCCCACCAAGGGGCCCATGGGCCCCC	276
Qy	383	C-----CTGGCTATGGGGGAAACCTTTCACTCCGACCTGGCCCTGGCCAGTCAAGGATGG	437
Db	277	CGGGCTCCCGGTACATGGGCAGCCCCGCGTGGACCCCGGCTGGCCCCCGC--GGATGG	335
Qy	438	ATCAGTCCCGCAAGAGACCTGCCTCTCAG-----CAGATCCAGCAGGT	480
Db	336	AGCCCGCCCGCAAGCAGACGCGCCCGCCGCGCAGAGCCAGGCACAGACCGCAGGGCCA	395
Qy	481	CCAGCAGAGCGCGTCCAAATCGAAACCAACATGCAAAAGAAAGAGATGCTGCACAA	540
Db	396	GCGGAGCCCAACGCCGCCCGCGGAGCCGAGTGCCCAAGAGGAGGAAGATGGCTGCACAA	455
Qy	541	AATTTCTACCTCAAAAGGATTCGTGAATCGTTACCAAGAATCCAGAGCTATATGGATCTCTT	600
Db	456	AATCCTCCCTCAAGGATTCGGGAGCTGGTCCCGAGTCCAGGCTTACATGGACCTCTT	515
Qy	601	GGCTTTTGAAGGAAACTGGACCAAGCTATCATGAGGAAAACGGCTAGATATCCAAAGAGCC	660
Db	516	GGCATTTTGAGAGGAAATCGGATCAACCAATCATCGGAAGGGGTGGACATCCAGGAGGC	575
Qy	661	CTTGAAACGTCCCATTAAGCAAAAACGGAAGCTCGGAATTTTCATTTCTTAACACTTTCAA	720
Db	576	TCTGAAGAGGCCCATGAAGCAAAAGCGGAAGTCCGACTCTATATCTCCAACTTTTTAA	635
Qy	721	TCCGGCTTAAGTCAGATCCGAGGATGGGAAGGACGGTGGCTTCCTGGGAGCTTCGGGT	780
Db	636	CCCTGCGAAGTCTGATGCTGAGGATTCGACGGCAGCATTCCTCTCTGGGAGCTACGGGT	695
Qy	781	AGAAGGACGGCTCTCGGAGGATTCAGGCTTGTCCAAATATGATGCCACTAAACAAAAAGAG	840
Db	696	GGAGGGAAAGCTCTCG-----ATGATCCACAGCAACAGAAAGCG	734
Qy	841	GAAGTCTCTCTCTCTTTTAAGTCTTGGTGATTGAATCTGCAAAAGACCTGTATGGGCC	900
Db	735	GAAGTCTCTCTCTTTCTTCAAGAGTTTGGTCAATCGAGCTGGACAAAAGATCTTTATGGCCC	794
Qy	901	AGACAAACATCTGTAGNATGGCACAGACCGCCACTACCCAGAGAGCCGATGGCTTCCA	960
Db	795	TGACAAACCACTCGTTGAGTGGCATCGGACACCCACAGCCAGAGACGGACGGCTTCCA	854
Qy	961	GGTGAAGCGCCAGGAGATGGAATGTACGGTGTACTGTCTACTGTAGTCTGGATTACCA	1020
Db	855	GGTGAACAGCCCTGGGACCTGAGTGTGCGCTGCACCGCTGCTCTCATGTCTGNACTACCA	914
Qy	1021	GCCTCCCCAGTTTAAATTTAGACCCCCCCTAGTCTCGACTCCTGGGCATCCATACCCAGAC	1080
Db	915	GCCTCCCCAGTTTCAAACTGGATCCCCGCCCTAGCCCGGCTGCTGGGGCTGCACACACAGAG	974
Qy	1081	TCGTCCAGTGNATCAACAGCACTGTGGCAATATATTAAAGACACATTAAGCTCCAGGACCC	1140
Db	975	CCGCTCAGACCATTTCCAGGCCCCCTGTGGCAGTATGTGAAGACCAACAGGCTGCAGGACT	1034
Qy	1141	TCAGAGCGGGAGTTTGTCTCTGTGACAGTACTCTGCAGCAGATCTTTTGAGACTCAACG	1200
Db	1035	CCATGACAAAGGAATACATCAATGGGACAAAGTATTTTCAGCAGATTTTGTGTTGCCCG	1094
Qy	1201	TATGAAGTTTTTCAGAGATCCCTCAGCGGCTCCATGCGCTTGGCTTATGCAACAGAACTTAT	1260

Qy 2650 AAAGCTTCAGACCCCTCAGGTAGCAGAGACCTTGATCTTGSCCCTTGATCTGAG 2709
|||
Db 171 AAAGCTTCAGACCCCTCAGGTAGCAGAGACCTTGATCTTGSCCCTTGATCTGAG 112
|||
Qy 2710 ATGGTTTTCGATCTTTCCAGAGAGCTCACAATTCTTCCAGGTGTATCACCCTCG 2769
|||
Db 111 ATGGTTTTCGATCTTTCCAGAGAGCTCACAATTCTTCCAGGTGTATCACCCTCG 52
|||
Qy 2770 AGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAGGTTGGAGACAG 2820
|||
Db 51 AGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAGGTTGGAGACAG 1
|||

Search completed: October 11, 2003, 14:36:56
Job time : 6896 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 11:17:57 ; Search time 517 Seconds
(without alignments)
14833.856 Million cell updates/sec

Title: U66617
Perfect score: 2841
Sequence: 1 GAATTCGGCTATCCCATG.....TGGGCAACAAGGGGATTC 2841

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: N.Geneseq_19jun03.*
 - 2: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
 - 3: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
 - 4: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
 - 5: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
 - 6: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
 - 7: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
 - 8: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
 - 9: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
 - 10: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
 - 11: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
 - 12: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
 - 13: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
 - 14: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
 - 15: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
 - 16: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
 - 17: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
 - 18: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
 - 19: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
 - 20: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
 - 21: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
 - 22: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
 - 23: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001.DAT.*
 - 24: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*
 - 25: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1457.4	51.3	2017	21	AAFP21750 Human breast and o
2	1330.4	46.8	3103	25	AB199242 Mouse ischaemic co
3	537.2	18.9	1452	24	AAD53119 Human BRG1-associa
4	488.8	17.2	606	24	ABK44460 cDNA encoding colo
5	425.8	15.0	2511	24	AB573872 Human cDNA encodin
6	425.4	15.0	3089	22	AAH98221 Human EST-derived
7	425.4	15.0	3272	22	AAH98319 Human EST-derived
8	414.2	14.6	2010	23	ABL02583 Drosophila melanog

9	414.2	14.6	4138	23	ABL02582 Drosophila melanog
C 10	408.2	14.4	6809	25	AB235936 Human secretory po
C 11	397	14.0	1970	24	ABO54477 Human ovarian anti
C 12	388.8	13.7	452	21	AAZ80712 Human colon cancer
C 13	310.2	10.9	690	21	AAZ80561 Human colon cancer
C 14	298.6	10.5	2581	25	ABX34564 Human mdt cDNA SE
C 15	221.6	7.8	755	25	AAD53120 Human BRG1-associa
C 16	188	6.6	5938	24	ABN80218 Human chemically m
C 17	145	5.1	5938	24	ABN80219 Human chemically m
C 18	123.4	4.3	310	22	AAF79889 Nucleotide sequenc
C 19	113.6	4.0	285	25	ABX47225 Bovine EST associa
C 20	96.6	3.4	95223	21	AAF22282 BAC containing rep
C 21	87	3.1	3824	23	ABL03782 Drosophila melanog
C 22	82.6	2.9	15500	22	ABAI5840 Human nervous syst
C 23	82.6	2.9	15500	22	AA336609 Human cardiovascular
C 24	82.6	2.9	15500	25	AB273784 Secreted protein g
C 25	82.6	2.9	15500	25	ABT16889 Human secreted pro
C 26	82.6	2.9	15500	25	ABZ67378 Human secreted pro
C 27	79.6	2.8	740	24	ABO65483 Arabidopsis thalia
C 28	76	2.7	450	24	ABT07117 Human ovarian can
C 29	76	2.7	450	25	ABX72995 Human ovarian carc
C 30	75	2.6	8788	22	AAI35930 Human musculoskele
C 31	75	2.6	8788	22	AAI62919 Human genomic DNA
C 32	75	2.6	8788	25	ABX58918 cDNA encoding nove
C 33	72.8	2.6	520	22	ABA29582 Probe #8048 for ge
C 34	72.8	2.6	520	22	AAK10561 Human brain expres
C 35	72.8	2.6	520	22	AAI17316 Probe #7749 for ge
C 36	72.8	2.6	520	22	AAI42206 Probe #10892 used
C 37	72.8	2.6	520	24	ABSI0454 Human genome-deriv
C 38	69.6	2.4	283	25	ABX85078 Corn ear-derived p
C 39	60.2	2.1	65	24	ABN51678 Human spliced tran
C 40	60	2.1	60	24	ABN33436 Breast cancer mark
C 41	57.8	2.0	335	25	ABT21717 Oligonucleotide fo
C 42	54.6	1.9	687	24	ABQ18226 Oligonucleotide fo
C 43	54.6	1.9	687	24	ABQ18227 Apeeryillus fumiga
C 44	51.6	1.8	1437	25	ABT20838 Apeeryillus fumiga
C 45	51.6	1.8	1437	25	ABT20838 Apeeryillus fumiga

ALIGNMENTS

RESULT 1	
ID	AAFP21750 standard; DNA; 2017 BP.
AAFP21750	
AC	AAFP21750;
XX	
DT	27-MAR-2001 (first entry)
DE	Human breast and ovarian cancer associated antigen gene SEQ ID 137.
XX	
XX	Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW	nocrotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW	antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
KW	antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW	Addison's disease; allergy; autoimmune haemolytic anaemia;
KW	multiple myeloma; rheumatoid arthritis; Crohn's disease;
KW	cardiovascular disorder; wound healing; neurological disease; ds.
OS	Homo sapiens.
XX	
PN	WO20055173-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US05881.
XX	
PR	12-MAR-1999; 99US-0124270.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	

PI Rosen CA, Ruben SM;
XX WPI; 2000-611515/58.
DR P-PSDB; AAB58847.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 1; Page 583; 1299pp; English.
XX
CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB59711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; anti-inflammatory; anti-ulcer; vulnery; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 2017 BP; 428 A; 589 C; 455 G; 536 T; 9 other;

Query Match 51.3%; Score 1457.4; DB 21; Length 2017;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1548; Conservative 8; Mismatches 10; Indels 10; Gaps 8;

QY 1272 ATGTCATCAGTGTTCACCCGAATGATCAGAAAGACACAGCTGTGTTATCAGATTGATGTTG 1331
DB 4 ATGTCATCAGTGTTCACCCGAATGATCAGAAAGACACAGCTGTGTTATCAGATTGATGTTG 63

QY 1332 AAGTGGATGACACCTTGAAGACCCAGATGAAATTTCTGCTGTCCACTGCCAGCCAAAC 1391
DB 64 AAGTGGATGACACCTTGAAGACCCAGATGAAATTTCTGCTGTCCACTGCCAGCCAAAC 123

QY 1392 AGGAGATTGCTACTCTAGACAAACAGACAAATGACTGATGTGGTGAACCCAGAGAGG 1451
DB 124 AGGAGATTGCTACTCTAGACAAACAGACAAATGACTGATGTGGTGGGTAA - CCARARGARG 182

QY 1452 AGCGCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGAGGCTGTGTGCCGATCTTCT 1511
DB 183 AGCGCGAGCTGAGTTCTACTTCCAGCCCTGGG - KCAGAGGCTGTGTGCCRATCTTCT 241

QY 1512 ACTCCAAGTGTGAGCAGAGACGACAAAGAAATAGACAAAGCCCTGGGAATCCGGAAATACAT 1571
DB 242 ACTCCAAGTGTGAGCAGAGACGACAAAGAAATAGACAAAGCCCTGGGAATCCGGNATACAT 301

QY 1572 AGGGCTCTCCACAGCCCTGATTCGACTGACACCAATTCGATTTGGGCCCTGTGCTGC 1631
DB 302 AGGGCTCTCCACAGCCCTGATTCGACTGACACCAATTCGATTTGGGCCCTGTGCTGC 361

QY 1632 CTGCTCATAGTATCTGCTTGGTCTGCTTGGGGCGTTCAGAGGAGTGTGTTGGTTCA 1691
DB 362 CTGCTCATAGTATCTGCTTGGTCTGCTTGGGGCGTTCAGAGGAGTGTGTTGGTTCA 421

QY 1692 AGGCAAGACCAAGATGAAGAGGGTCTCACAAGACACCTGTTATCTCTTCTTTCCACCT 1751
DB 422 AGGCAAGACCAAGATGAAGAGGGTCTCACAAGACACCTGTTATCTCTTCTTTCCACCT 481

QY 1752 ATCTCTTCCACCCAGCTTCCCTTTGGCCCAACAAAGTTCCAGTGTGCTGTACCTTCC 1811
DB 482 ATCTCTTCCACCCAGCTTCCCTTTGGCCCAACAAAGTTCCAGTGTGCTGTACCTTCC 541

QY 1812 CCTGGTCTCATATAGGACCTCTAGATAGTGTTAGAGAGAGAACATGTAGTGGTAATGAGTG 1871
DB 542 CCTGGTCTCATATAGGACCTCTAGATAGTGTTAGAGAGAGAACATGTAGTGGTAATGAGTG 601

QY 1872 CTTGGAATGGAATT - GGCTCAGGCCAGGTGGTCTTCAAGGGGACACAGCTAATCATCTTA 1930
DB 602 CTTGGAATGGAATTGGGCTCAGGCCAGGTGGTCTTCAAGGGGACACAGCTAATCATCTTG 661

QY 1931 CCCTTCAGAGACCCAGGAGTTGGG - - -TTTCGGTCCCTTCTCCAAAGACTCAGGCGCTGGGG 1987
DB 662 CCCTTCAGAGACCCAGGAGTTGGAGCTTTCGGCTCCTTCTCCAAAGACTCAGGCGCTGGGG 721

QY 1988 CACTCTATAAGCTAGTGTGATCTTGGCTCTCTCTGATAACAGAAATCCAAATTTCTTCTTCC 2047
DB 722 CACTCTATAAGCTAGTGTGATCTTGGCTCTCTCTGATAACAGAAATCCAAATTTCTTCTTCC 781

QY 2048 CTCACAGGTTTGGAAACAAACTCTCCCTCAGTGTGTCCTGTAGCACTACAGAAACCC 2107
DB 782 CTCACAGGTTTGGAAACAAACTCTCCCTCAGTGTGTCCTGTAGCACTACAGAAACCC 841

QY 2108 TGGTTCCT - GGCTCCACTGAGCCCCAGTCACTGCCAG - CCTCTGGGTTGGCCTGCTGT 2165
DB 842 TGGTTCCTTGGGCTCCACTGCCCCAGGGTCACTGCCAGCCCTCTGGGTTGRCCTGCTGT 901

QY 2166 CAGTGTCTTCTCTCACTCCTTAGTTGGGGTCCACATCAGTATTTGGAGTTTGTCTTTATT 2225
DB 902 CAGTGTCTTCTCTCACTCCTTAGTTGGGGTCCACATCAGTATTTGGAGTTTGTCTTTATT 961

QY 2226 GCTCCCTCCACAGACATCCCTGTGGTGGCTTGTGATTCCTCAGATCGCCCTAATC 2285
DB 962 GCTCCCTCCACAGACATCCCTGTGGTGGCTTGTGATTCCTCAGATCGCCCTAATC 1021

QY 2286 CCGGGCATTTGGGTGGGGGAATCTTGGCTTTCCTTTCCAGAGCCCGAGGATCTCATCTG 2345
DB 1022 CCGGGCATTTGGGTGGGGGAATCTTGGCTTTCCTTTCCAGAGCCCGAGGATCTCATCTG 1081

QY 2346 GGGAACTGTCTATTCGCCAGCAGAGGCTGTTCCTTCTCAGTGTGGAGATGTGACTCATTC 2405
DB 1082 GGGAACTGTCTATTCGCCAGCAGAGGCTGTTCCTTCTCAGTGTGGAGATGTGACTCATTC - 1140

QY 2406 CATTCACCTCATCTCACCTCGCTCTGCATCCCTTAATGGAGAAACCGGCGCTAAACCAAA 2465
DB 1141 CATTCACCTCATCTCACCTCGCTCTGCATCCCTTAATGGAGAAACCGGCGCTAAACCAAA 1200

QY 2466 CGGTAAGAAAGCCCTGGGCCATCCCTGTCTTCTGTCCCTTCTGCCAGTTGACACC 2525
DB 1201 CGGTA - AAAAAGCCCTGGGCCATCCCTGTCTTCTGTCCCTTGTCTGCCAGTTGACACC 1259

QY 2526 TACTGTGACTTCTTAGGGCACTGAGGAGTGAAGCGCTAGGGCTGGAGAAATAGCGCTGA 2585
DB 1260 TACTGTGACTTCTTAGGGCACTGAGGAGTGAAGCGCTAGGGCTGGAGAAATAGCGCTGA 1319

QY 2586 GTTGGGTTTGTGACTTCTCCCTCTCCCTGCCTCAGAGGATGTGACTCCCCAGCCCTGC 2645
DB 1320 GTTGGGTTTGTGACTTCTCCCTCTCCCTGCCTCAGAGGATGTGACTCCCCAGCCCTGC 1379

QY 2646 CCTCAAGGCTTCAGACCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGGCCCTTGGATC 2705
DB 1380 CCTCAAGGCTTCAGACCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGGCCCTTGGATC 1439

QY 2706 TGAGATGGTGTGCTTTCAGAGGAGGCTCACAATCTTCTTCCAGGTTGTATCACC 2765
DB 1440 TGAGATGGTGTGCTTTCAGAGGAGGCTCACAATCTTCTTCCAGGTTGTATCACC 1499

QY 2766 CCGGAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGGTGGGAGACAGCTGGG 2825
DB 1500 CCGGAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGGTGGGAGACAGCTGGG 1559

QY 2826 CACAAGGGGGATTC 2841
DB 1560 CACAAGGGGGAAATTC 1575

```

RESULT 2
ABI99242
ID ABI99242 standard; cDNA; 3103 BP.
XX
AC ABI99242;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:65.
XX
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX   vasospastic ischaemia; ischaemic condition; ischaemic disease; sg.
XX
OS Mus musculus.
XX
PN W0200188188-A2.
XX
PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001MO-JP04192.
XX
PK 18-MAY-2000; 2000JP-0145977.
XX
PA (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Negata T, Ishii Y;
XX
DR WPI; 2002-034733/04.
XX
PS P-PSDB; ABB57044.
XX
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes.
XX
PS Claim 2; Page 203-208; 2690pp; English.
XX
CC The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (1) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (1). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischaemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ Sequence 3103 BP; 657 A; 925 C; 817 G; 704 T; 0 other;
XX
Query Match 46.8%; Score 1330.4; DB 24; Length 3103;
Best Local Similarity 75.3%; Pred. No. 0;
Matches 2124; Conservative 0; Mismatches 431; Indels 265; Gaps 27;
XX
QY 158 CCGGCGGGTTTTCAGTCTGTGGCTCAAGCGGCGCGCGAGCTTCAGAGAGGCGGG 217
Db 1 CCGGCGGGTTTTCAGTCTGTGGCTCAAGCGGCTCCCGGAGGCTTCAGAGAGGCGGG 60
QY 218 CCGGCGTGTCTGTGGGCCCGGCGGAATCTCCGGGCTTCCTGTGGAAATGGGCCGGCT 277
Db 61 CCGGCGGCTGTCTGTGGGCCCGGCGGAATCTCCGGGCTTCCTGTGGAAATGGGCCGGCG 120
QY 278 CCGGCGTCAAGGCTGATCCGCTCCCGATGCCGCGAGCGGCTATCCGAGACAGATAG 337
Db 121 CCGGCGTCAAGGCTGATCCGCTCCCGATGCCGCGAGCGGCTATCCGAGACAGATAG 180
QY 338 TTGCGAGGAGCCGATGACACTTCAGGAGCTTTCATGAGGAGCCCTGGCTATGCGGGG 397
Db 1178 CAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGATCCCTCAGGGGCTCATGCC 1237
Db 1021 CAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGATCCCTCAGGGGCTCATGCC 1080
QY 1238 TTGCTTATGCAACGAGACCTATCATCTATATCATGTCATCATGATGACCCGATGAT 1297
Db 1081 TTGCTTATGCAACGAGACCTATCATCTATATCATGTCATCATGATGACCCGATGATG 1140
QY 1298 CAGAAAAAGACAGCTTTGTTATGACATTTATGATTTGATGATGATGACCTTGAAGACCGAG 1357
Db 1141 CAGAAAAAGACAGCTTTGTTATGACATTTATGATTTGATGATGATGACCTTGAAGACCGAG 1200
QY 1358 ATGAAATCTTTTCTGCTGTCATGTCAGCCAGCAAGAGAGATTTGACTCTAGACAA 1413
Db 1201 ATGAAATCTTTTCTGCTGTCATGTCAGCCAGCAAGAGATTTGACTCTAGACAAAG 1260
QY 1414 ----- 1413
Db 1261 ATTCATGAGATGATGAGACCATCAACGAGTGAAGACCGAGAGATTATGTTGAGC 1320

```

```

Db 181 CTGCCAGGTAGCCGATGACACCTTCAGGGACCTTCCTCATGGACCTTCGCTATGCGGGG 240
QY 398 AACCTTCAATCCGACCTGGGCTGGCCAGTCCAGGATGATCATCCCGCAAGAGACCT 457
Db 241 AACCTTCAATCCGACCTGGGCTGGCCAGTCCAGGATGATCATCCCGCAAGAGACCT 300
QY 458 GCCCTCAGCAGATCCAGAGGTCAGAGCGAGCGGTCCAAATCGAAATCGAAATCGAATGCA 517
Db 301 GCACCTCAACAGATCCAGAGGTCAGAGCGAGCGGTCCAAATCGAAATCGAATCGAATGCA 360
QY 518 AAGAAAAAGAGATGGCTGACAAATTTCTACTCAAGAGATTCGTGAATCTGTATCCAGAA 577
Db 361 AAGAAAAAGAGATGGCTGACAAATTTCTACTCAAGAGATTCGTGAATCTGTATCCAGAA 420
QY 578 TCCGAGGCTATATGATCTCTTGGCTTTTGAAGGAAATCGAGACCTATCATGAGG 637
Db 421 TCACAGGCTTACATGATCTCTTGGCTTTTGAAGGAAATCGAGACCTATCATGAGG 480
QY 638 AACGCGTATGATATCCAGAGGCTTTGAACCTCCATTAAACCAAAAACGAAAGCTGCGA 697
Db 481 AAGCGGCTATGATATCCAGAGGCTTTGAACCTCCATTAAACCAAAAACGAAAGCTGCGA 540
QY 698 ATTTTCATTTCTAACAATTTCAATCCGCTAAATGATGATCCGAGATGCGGAGAGGAGC 757
Db 541 ATTTTCATTTCTAACAATTTCAATCCGCTAAATGATGATGATGATGAGGAGAGGAGC 600
QY 758 GTGGCTTCTGGGACCTTCGGGTAGAGAGCGGCTCTCGAGGATTCAGCCTTTCCTCAA 817
Db 601 GTGGCTTCTGGGACCTTCGGGTAGAGAGCGGCTCTCGAGGAGCGGCTTCCTTCCTCAA 660
QY 818 TATGATGCGACATTAACAAAGAGAGGATTTCTCTCTTCTTAAGTCTCTGATGATGAA 877
Db 661 TATGATGCGACATTAACAAAGAGAGGATTTCTCTCTTCTTAAGTCTCTGATGATGAA 720
QY 878 CTGGAACAAAGACTGATAGGCGCAGACCAATCTGTAGAAATGACACAGACCGCCACT 937
Db 721 CTGGAACAAAGACTGATAGGCGCAGACCAATCTGTAGAAATGACACAGACCGCCACT 780
QY 938 ACCGAGAGACCGATAGGCTTCCAGGTGAAGCGGCCAGAGATGTGAATGTACGCTACT 997
Db 781 ACCGAGAGACCGATAGGCTTCCAGGTGAAGCGGCCAGAGATGTGAATGTACGCTACT 840
QY 998 GTCTACTGATGCTGATTTACAGGCTCCCGAGTTTAATAGACCCCGCTAGTCTGCA 1057
Db 841 GTCTACTGATGCTGATTTACAGGCTCCCGAGTTTAATAGACCCCGCTAGTCTGCA 900
QY 1058 CTCTGGGCAATCATACCCAGACTGTCAGTATGATTCAGCACTGTGGCAATATAT 1117
Db 901 CTCTGGGCAATCATACCCAGACTGTCAGTATGATTCAGCACTGTGGCAATATAT 960
QY 1118 AAGACATTAAGCTTCAGAGACCTTACAGAGCGGAGTTTTCATCTGTGACAAATACCTG 1177
Db 961 AAGACATTAAGCTTCAGAGACCTTACAGAGCGGAGTTTTCATCTGTGACAAATACCTG 1020
QY 1178 CAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGATCCCTCAGGGGCTCATGCC 1237
Db 1021 CAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGATCCCTCAGGGGCTCATGCC 1080
QY 1238 TTGCTTATGCAACGAGACCTATCATCTATATCATGTCATCATGATGACCCGATGAT 1297
Db 1081 TTGCTTATGCAACGAGACCTATCATCTATATCATGTCATCATGATGACCCGATGATG 1140
QY 1298 CAGAAAAAGACAGCTTTGTTATGACATTTATGATTTGATGATGATGACCTTGAAGACCGAG 1357
Db 1141 CAGAAAAAGACAGCTTTGTTATGACATTTATGATTTGATGATGATGACCTTGAAGACCGAG 1200
QY 1358 ATGAAATCTTTTCTGCTGTCATGTCAGCCAGCAAGAGAGATTTGACTCTAGACAA 1413
Db 1201 ATGAAATCTTTTCTGCTGTCATGTCAGCCAGCAAGAGATTTGACTCTAGACAAAG 1260
QY 1414 ----- 1413
Db 1261 ATTCATGAGATGATGAGACCATCAACGAGTGAAGACCGAGAGATTATGTTGAGC 1320

```


XX Example 1; Column 55; 31bp; English.

CC The invention relates to human nuclear receptor cofactor polypeptides.
CC The invention particularly relates to proliferator-activated receptor
CC (PPAR)-gamma and retinoid X receptor (RXR) cofactor polypeptides.
CC Polypeptides of the invention are useful for screening compounds that
CC modulate the interaction of the nuclear receptor with the nuclear
CC receptor cofactor. Modulators of the invention are useful in preparing
CC a pharmaceutical composition for treating and/or preventing diseases
CC or pathologic conditions associated with cell types that express PPAR
CC receptors. The pathologic conditions treated include metabolic or cell
CC proliferative disorders such as diabetes, cardiovascular disorders
CC (e.g. atherosclerosis), renal diseases, neurodegenerative diseases
CC (e.g. Parkinson's disease, Alzheimer's disease), inflammatory diseases
CC (e.g. psoriasis, acne), wounds, osteoporosis, infections (e.g. HIV)
CC or cancer. The invention is useful in gene therapy and as vaccines.
CC The present sequence is human BRG1-associated factor (BAF) 6002 DNA.
CC This sequence is used in the exemplification of the invention.

XX Sequence 1452 BP; 330 A; 447 C; 424 G; 251 T; 0 other;

Query Match 18.9%; Score 537.2; DB 25; Length 1452;
Best Local Similarity 70.2%; Pred. No. 1.3e-137;
Matches 785; Conservative 0; Mismatches 288; Indels 45; Gaps 3;

QY 325 GAGACAGGTATGTCGACGAGCCGAAATGACACTTCCAGGACCTTCATGGAGACCC 384
DB 78 GCGCCCGGATGCTCTGAGACCGGATGCCCCACAGGGGGCGCCCATGGCCCCC 137
QY 385 TGGC-----TATGGGGGAAACCTTCACTCCAGCTGAGCCCACTCAGATGGA 438
DB 138 GGGCTCCCGTATCATGGGAGACCCCGCGTGCACCCGCGCTGGCCCCCGGCGATGA 197
QY 439 TCAATCCGCAAGAGACTGCTCCCTCAG-----CAGATCCAGAGGT 480
DB 198 GCGCGCCCGCAAGCGAGCGCCCGCGGAGAGCCAGCGCAGAGCGAGGCGCA 257
QY 481 CCAGCAGCAGGGGCTCCAAATGAAACCAATGCAAGAAAGAAAGATGGCTACAA 540
DB 258 GCGGTCGCCACCGCCCGCGGAGCGCAGTGCAGAGAGAGAAATGGCTACAA 317
QY 541 AATTCTACCTCAAAAGATTGTAACCTGATCCAGATCCAGGCTTATGATCTCT 600
DB 318 AATCTCCCTCAAGAGATTGGAGCTGTGCTCCGAGCTCCAGGCTTACATGACCTCTT 377
QY 601 GGGCTTTGAAAGAACTGACCCAGATCTATCATGAGAAACGGCTGATATCCAGAGGC 660
DB 378 GGCATTTGAGAGAACTGATCAACCATATGCGAAGCGGCTGATCATCCAGAGGC 437
QY 661 CTTGAAAGCTCCCATTAAGCAAAAGGAAAGCTGCAATTTTCACTTTCACTTTCA 720
DB 438 TCTGAAGAGCCCTATGACAAAGCGAAGCGCAGCTATATCTCCAAACCTTTTAA 497
QY 721 TCCGGCTAAGTACAGATGCGGAGATGGGGAAGGAGCGTCTCTGGAAGCTTGGGT 780
DB 498 CCTTGGAGAGCTGATGCTGAGATTCGACGACACATTCCTCTGGAAGTACGGGT 557
QY 781 AGAAGACGCGCTCTGAGAGATTCAAGCTTGTCAATATGATGCCACTAAACAAAGAG 840
DB 558 GAGGGGAAAGCTCTCG-----ATGATCCAGCAAGCAAGAGG 596
QY 841 GAAGTCTCTCTCTTTTAAGTCTTGTGATGATGACGAAAGACCTGTATGGCC 900
DB 597 GAAGTCTCTCTCTTTTCAAGAGTTGTCATGAGCTGAGCAAAAGATCTTTATGCC 656
QY 901 AGACACCACTCTGTGATGAGCAGACCGCACTACCCAGGAGACCGATGGCTTCA 960
DB 657 TGAACAACCACTCTGTGATGAGCAGACCGCACTACCCAGGAGACCGCTTCA 716
QY 961 GGTGAAGCGGCGAGAGATGTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 717 GGTGAAGCGGCGGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 776

QY 1021 GCGTCCCGAGTTTAATTAGACCCCGCTAGCTGACTCTGGGATCCATACCCAGAC 1080
DB 777 GCGTCCCGAGTTTAATTAGACCCCGCTAGCTGACTCTGGGATCCATACCCAGAC 836
QY 1081 TCGTCCAGTATCATCAGACAGCTGTGGCAATATATTAAGACATAAGCTCCAGACC 1140
DB 837 CCGTCCAGTATTCAGAGCCCTGTGGCAGATATGAAGACCAAGGCTCCAGAGCTC 896
QY 1141 TCACAGAGGAGTTTTCATCTGTGACAGACTCTGACAGATCTTGAAGCTCAACG 1200
DB 897 CCATACAAAGGATATCATCATATGAGGAGACAGATTTTCAGACAGATTTTGAATG 956
QY 1201 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCATGCTTGTATGACACAGAACTAT 1260
DB 957 GGTGAAGTTTTCAGAGATCCCTCAGCGGCTCATGCTTGTATGACACAGAACTAT 1016
QY 1261 CATCATTAATCATGTCATCATGTTGACCCGATGATCAGAAAAGACAGCTTGTATGA 1320
DB 1017 TGTCAATCAACATGTCATCATGAGGTGACCTTCAGACCAAGAAAGAGCGGCTATGA 1076
QY 1321 CATTAATGTAAGTGATGACACCTTGAAGACCAAGATGATTTCTTCTGCTCCAC 1380
DB 1077 CATTAATGTAAGTGATGACACCTTGAAGACCAAGATGATTTCTTCTGCTCCAC 1136
QY 1381 TGCAGCCCAAGAGATGTTGCTACTCTAGACAAACAGA 1418
DB 1137 GGCACCAAGCAGAGATCATGCTCTTGGACATGAGA 1174

RESULT 4
ID ABK4460/C
ID ABK4460 standard; cDNA; 606 BP.
XX ABK4460;
XX 05-JUN-2002 (first entry)
XX cDNA encoding colon tumour protein, SEQ ID No 11.
XX Human; colon tumour; vaccine; colon cancer; immunogenic;
XX Immunotherapy; gene; ss.
XX Homo sapiens.
XX WO200212328-A2.
XX 14-FEB-2002.
XX 31-JUL-2001; 2001WO-US24218.
XX 03-AUG-2000; 2000US-223283P.
XX 28-MAR-2001; 2001US-279763P.
XX 29-JUN-2001; 2001US-302051P.
XX (CORI-) CORIXA CORP.
XX King GE, Meagher MJ, Xu J, Secretist H;
XX WPI; 2002-241739/29.
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
XX for diagnosing, preventing, and treating colon cancer, and as markers
XX for the progression of cancer -
XX Claim 1; SEQ ID No 11; 147bp; English.
XX The invention relates to polynucleotides encoding colon tumour proteins.
XX The polynucleotides and encoded polypeptides are useful in pharmaceutical
XX compositions, such as vaccines, for the diagnosis, prevention, and
XX treatment of colon cancer. Polynucleotide sequences may be used as
XX hybridisation probes or primers, and in the design and preparation of
XX ribozyme molecules for inhibiting expression of tumour polypeptides and

CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: with the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.

XX
SQ Sequence 606 BP; 151 A; 148 C; 184 G; 118 T; 5 other;

Query Match 17.2%; Score 488.8; DB 24; Length 606;
Best Local Similarity 93.6%; Pred. No. 2e-124;
Matches 553; Conservative 0; Mismatches 30; Indels 8; Gaps 4;
QY 2237 GACATCCCTGTGGTGGCC---TTTGATTCCTCAGATCGCC---TAATCCCGGG 2290
Db 590 GACACTCCCTGTGGTGGCCCTTTGTGAATCCCTTCAGATMTGCCCTAATCCCGGG 531

QY 2291 CATTTGGGTGGGGGAATCTTGCTTTCCTTTTCAGAGCCCGAGGATCTCATCTGGGAA 2350
Db 530 CATTTGGGGGGGAATCTTGCCNTTCCCTTCAGAGCCCGAGGATCTCATCTGGGAA 471

QY 2351 CTGTCATTTGCCAGCAGAGCGTGTTCCTTCCTGAGTT-TGAGAGTGTGACTCATTTCAATT 2409
Db 470 CTGTCATTTGCCAGCAGAGCGTGTTCCTTCCTGCTGTTTGTGAGATGTGACTCATTTCAATC 411

QY 2410 CACTCAGTCCACCTCGCTCTGCATCCCTTAATGGAGAAAGGGCTAAACCCAAACGGG 2469
Db 410 ACTCATCCACCTCGCTCTGCATCCCTTAATGGAGAAAGGGCTAAACCCAAACGGG 351

QY 2470 TAAAAAGCCCTGGGCCATCCCTGTCTTCTGCTTCTGCTCCAGTGTGACACCTACT 2529
Db 350 T-AAAAAGCCCTGGGCCATCCCTGTCTTCTGCTTCTGCTCCAGTGTGACACCTACT 292

QY 2530 GGTGACTTCTAGGCACTGAGGAGTGAAGCGCTTAGGCTGGAGATAGCGCTGAGTTG 2589
Db 291 GGTGACTTCTAGGCACTGAGGAGTGAAGCGCTTAGGCTAAAGATAGCGCTGAGTTG 232

QY 2590 GGTGTTGTGACTTCCCTCTCCCTGCTCAGAGATTTGACTCCCGAGCCCTGCCCTC 2649
Db 231 GGTGTTGTGACTTCCCTCTCCCTGCTCAGAGATTTGACTCCCGAGCCCTGCCCTC 172

QY 2650 AAAGCTTCAGACCCCTCAGGTAGCAGGAGCCTTGTGATCTTGCCCTTGGATCTGAG 2709
Db 171 AAAGCTTCAGACCCCTCAGGTAGCAGGAGCCTTGTGATCTTGCCCTTGGATCTGAG 112

QY 2710 ATGTTTTTGTGATCTTTCCAGGAGCCTTCAATCTTCTTCCAGGTTGTATCAACCCCG 2769
Db 111 ATGTTTTTGTGATCTTTCCAGGAGCCTTCAATCTTCTTCCAGGTTGTATCAACCCCG 52

QY 2770 AGTTAGCATATCCAGGCTCGAGACTCAACACAGCAAGGTTGGAGACAG 2820
Db 51 AGTTAGCATATCCAGGCTCGAGACTCAACACAGCAAGGTTGGAGACAG 1

RESULT 5
ABS73872

ID ABS73872 standard; cDNA; 2511 BP.

XX

XX ABS73872;

DT 06-DEC-2002 (first entry)

XX Human cDNA encoding NAAP12, Incyte 31123900CB1.

XX Human; ss; gene; nucleic acid-associated protein; NAAP; gene therapy;

KW microarray; proliferative disorder; developmental disorder;

KW cardiovascular disorder; neurological disorder; autoimmune disorder;

KW inflammatory disorder; atherosclerosis; bursitis; cirrhosis; hepatitis;

KW psoriasis; cancer; lymphoma; melanoma; brain cancer; breast cancer;

KW Cushing's syndrome; Alzheimer's disease; AIDS; Creutzfeldt-Jakob disease;

KW acquired immunodeficiency syndrome; Goodpasture's syndrome;

Crohn's disease; multiple sclerosis.

Homo sapiens.

W0200274913-A2.

XX 26-SEP-2002.

XX 14-MAR-2002; 2002WO-US07869.

XX 16-MAR-2001; 2001US-276857P.

PR 19-APR-2001; 2001US-285489P.

PR 04-MAY-2001; 2001US-285566P.

PR 04-MAY-2001; 2001US-288646P.

PR 11-MAY-2001; 2001US-288700P.

PR 11-MAY-2001; 2001US-290510P.

PR 16-NOV-2001; 2001US-290369P.

XX 16-NOV-2001; 2001US-332426P.

(INCY-) INCYTE GENOMICS INC.

Hillman JL, Baughn MR, Swarnakar A, Yue H, Elliott VS, Burford N;

Ding L, Tang YT, Lee SY, Azimzai Y, Wallia NK, Gietzen KJ;

Griffin JA, Lal PG, Yang J, Borowsky ML, Richardson IW, Yue H;

Becha S, Forsythe IJ, Jones KA, Warren BA, Thangavelu K;

Honchell CD, Jolley HE, Hafalia AJA, Ring HZ;

WPI: 2002-723448/78.

P-PSDB; ABG95655.

New human nucleic acid-associated proteins and polynucleotides, useful

for diagnosing, treating or preventing cardiovascular, neurological,

autoimmune or inflammatory disorders, e.g. atherosclerosis, Alzheimer's

or AIDS

Claim 5; Page 205-206; 219pp; English.

The invention relates to an isolated human nucleic acid-associated

protein (NAAP), comprising the protein sequences appearing as ABG95644-

ABG95669, or an immunogenic or biologically active fragment. Also

included are the NAAP encoding polynucleotides, a recombinant

polynucleotide comprising a promoter sequence operably linked to the

NAAP polynucleotide, a cell transformed with the recombinant NAAP

polynucleotide, an NAAP transgenic organism, an anti-NAAP antibody,

screening compounds as modulators or ant/agonists of NAAP or

modulators of NAAP polynucleotide expression, a microarray comprising

30-60 nucleotides of the NAAP polynucleotides and generating an

expression profile of a sample that contains NAAP polynucleotides.

The polypeptide, polynucleotide and agonist are useful for

treating a disease or condition associated with decreased expression of

functional NAAP. The antagonist is useful for treating a disease or

condition associated with overexpression of functional NAAP. The

antibody that specifically binds to the polypeptide is useful for

diagnosing a condition or disease associated with the expression of

NAAP. These diseases include proliferative, developmental,

cardiovascular, neurological, or autoimmune or inflammatory disorders.

These polypeptides, polynucleotides, agonists and antagonists may also

be used for preventing these diseases. These disease or conditions

associated with functional NAAP expression also includes

atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis, cancers

(e.g. lymphoma, melanoma, brain cancer or breast cancer), Cushing's

syndrome, Alzheimer's disease, acquired immunodeficiency syndrome (AIDS),

Creutzfeldt-Jakob disease, Goodpasture's syndrome, Crohn's disease, or

multiple sclerosis and many other diseases and conditions listed in

the specification. The present sequence encodes an NAAP protein.

Sequence 2511 BP; 512 A; 713 C; 760 G; 526 T; 0 other;

Query Match 15.0%; Score 425.8; DB 24; Length 2511;

Best Local Similarity 63.7%; Pred. No. 1.1e-106;

Matches 728; Conservative 0; Mismatches 382; Indels 33; Gaps 4;

QY 297 GCTCCCCGATGCCCGAGCGCGCTTATCCGAGACAGGTATGTTGCCAGGAGCCGAATGA 356

```

Db      267  GCGCCGCGGGGCCCCGGGGGCGCATACGAGGACCTGTCATGTCAACGAGGAAACCGGATGC 326
Qy      357  CACCTTAGGGACCTTCCATGGAGCCCCCTG-----GCTATGGGGGAAACCTTTCAG 407
Db      327  CCATGGCTGGCTTGCAGGTGGGACCCCTGCTGGCTCCCATTTGGTGAAGCAGCTCCGC 386
Qy      408  TCCGACCTGGCCGCGCCGATCAGGATGGATCAGTCCCGCAAGACAGACCTGCCCTCAGC 467
Db      387  TTGCACTTGGCATGCCACCCATCATGTGATTCATTCGAAACGCTGCTTGTGCCCC 446
Qy      468  AGATCCAGAGGTCGACGACGACGCGTCCAAATCGAAACCAATGCAAAAGAAAAAGA 527
Db      447  AG-----GGGAGCCTCCCATGCTGCCAGCGCGGGGTTAAAGAGAGGA 494
Qy      528  AGATGGCTACAAAATTCTACCTCAAGATTCGTGAACCTGGTACCAAGATCCAGCGCT 587
Db      495  AGATGGCAATAGTTTCTACCTCAGCGAATCCGGAGGCTTGTTCAGAGTCTCAGGCGT 554
Qy      588  ATATGGATCTCTTGGCTTTTGAAGAAGTGGACAGACTATCATGAGGAAACGGCTAG 647
Db      555  ACATGGATCTCTTGGCTTTTGAAGAGTGGACAGACCATTCTCCGACGCGGATGC 614
Qy      648  ATATCCAGAAGGCTTGAACGTCCTCATTAAGCAAAAAGGAAAGCTGCAATTTTCATT 707
Db      615  AGATCCAGAGGCGCATCAAAAAGCCTCTGACACAAAAGGAAAGCTTCGATCTACATT 674
Qy      708  CTAACTTTCAATCCGGCTAAGTCAGATGCGGAGATGGGAGGAGCGTGGCTTCTCT 767
Db      675  CCAATACGTTCAGTCCAGCAAGCGGAAAGCGATAGTGCAGAACTGCAGAGGAGCCCTG 734
Qy      768  GGGAGCTTC-----GGGTAGAAAGAGCGCTCCGAGAGATTCAGCCTTGCTCCAA----- 817
Db      735  GGGGAAACCCGAGAGGAGGAGCAAGTGGCTTCTGGGAACTCCGAGTGAAGGAAACCTGC 794
Qy      818  --TATGTGCACTAAACAAAAGAGAAAGTCTCTTCTCTTCTTAAAGCTTGGTATTTG 875
Db      795  TGGATATCTTACGAAACAGAAAGAGAAAGTTTCTTCTTCTTAAAGAGCTGTGATTTG 854
Qy      876  AACTGCAAAAGACTGTATGGGCGCAGACCAACCATGTGTGAATGGCACAGGACCGCA 935
Db      855  AGCTGGACAAGAGCTGTACGGGCTGACAAATCACTGTGGAGTGGACCGGATCCCA 914
Qy      936  CTACCCAGAGACCGATGGCTTCCAGGTGAAGGCGCCAGAGATGTGAATGTACGTTGA 995
Db      915  CCACCCAGAGAGAGATGGCTTCCAAGTAAAGGCGCTGGAGACCTCAAGTCAAGTGA 974
Qy      996  CTGTCTACTGATGCTGGATTACAGAGCTCCCGAGTTTAAATTAGACCCCGGCTAGCTC 1055
Db      975  CCTCTCTGCTCATGCTGGATCATCAAGCTTCCCAAGTACAAATTGACCCCGCATTTGCA 1034
Qy      1056  GACTCTGGGCACTCATACCAAGACTGTCAGTGCATCATCAAGCACTGTGGCAATATA 1115
Db      1035  GGGTGTGGGAGTGACACCGCAAGAGGCGCCCATCATGACGGCCCTGTGGCTTTTACA 1094
Qy      1116  TTAAACACATAAGCTTCCAGGACCTTCAAGAGCGGAGTTGTTCATCTGTGACAAAGTAC 1175
Db      1095  TCAAGACACAACGAGCTGCAAGATGGGACAGAGGGAATGATCACTGCAACCGTTACT 1154
Qy      1176  TGCAGAGATCTTTGAGACTCAACGATATGAAGTTTTCAGAGATCCCTCAGCGGCTCATG 1235
Db      1155  TCCGCAAGATCTTCAAGTTGTGGCGGACTCCGTTTCTCCAGATTTCCATGAGAGCTGCA 1214
Qy      1236  CCTTGTATGCCACGAGAACTATCATCTTAATCATGTTCATCAGTGTGACCCGAAATG 1295
Db      1215  GGTGGTGAGAGATCCAGACCCCATGTGTATCAACATGTCATTAAGTGTGACCTTAAG 1274
Qy      1296  ATAGAAAAAGAGAGCTTGTATGACATTTATGTTGAATGGATGACACTTGAAGACCC 1355
Db      1275  ACCAGAAAGAGAGAGCTGTGATGACATCATGTGGAGGTGAGACCACTGAAGAGCC 1334
Qy      1356  AGATGATTTCTTTTGTGCTGTCACTGGCCAGCAAGAGAGATTTGACTGTAGACAA 1415

```

```

Db      1335  AATAGCAATTTTCTGCTCTTACCAACCAATCAGAGAGATGCTCCCTTGATGTCA 1394
Qy      1416  AGA 1418
Db      1395  AGA 1397

RESULT 6
AAH98221
ID AAH98221 standard, cDNA, 3089 BP.
XX
XX AAH98221;
AC
XX
DT 12-OCT-2001 (first entry)
XX
XX Human EST-derived coding sequence SEQ ID NO: 78.
DE
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; BS.
XX
XX Homo sapiens.
OS
XX WO200154477-A2.
PN
XX 02-AUG-2001.
PD
XX
XX 25-JAN-2001; 2001WO-US02687.
PF
XX
XX 25-JAN-2000; 2000US-0491404.
PR
XX 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Wehrman T;
PI
XX WPI; 2001-476164/51.
DR
XX P-PSDB; AAM23562.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
XX Claim 1; Page 231-232; 1275bp; English.
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
XX Sequence 3089 BP; 648 A; 954 C; 946 G; 541 T; 0 other;
SQ

Query Match 15.0%; Score 425.4; DB 22; Length 3089;
Best Local Similarity 72.6%; Pred. No. 1.5e-106;
Matches 579; Conservative 0; Mismatches 211; Indels 7; Gaps 2;

```

Db 616 GAAGCAAAAGCGAAGCTGCGACTCTATATCTCCAAACATTTTAAACCTCGGAAGCCCTGA 675
Qy 736 TGCCGAGGATGGGAAGGACGGTGGCTTCTGGGAGCTTCGGGTAGAAGACCGCTCCT 795
Db 676 TGCTGAGGATCCGACGCGAGCATTTGCTCTGGGAGCTACGGGTGGAGGGAAGCTCCT 735
Qy 796 GGAGGATTACGCTTTGTCCAAATATGATGCCACTAAACAAAAGAGGAAGTTCTCTTCCTT 855
Db 736 GGATGAT-----GTACGTCCCGGCCCGCCAGCAGCAACAGAAAGCGGAAGTTCTCTCTTT 790
Qy 856 CTTTAAGTCTTTGTTGATTTGAATCTGGACAAAGACCTGTATGGCCAGACACACCATCTGCT 915
Db 791 CTTCAAGAGTTTGGTTCATCGAGCTGACAAAGATCTTTATGGCCCTGCAACACCACTCGT 850
Qy 916 AGAATGGCACAGGACCGGCTACCCAGGACGCGATGGCTTCCAGGTGAAGCGCCAGG 975
Db 851 TGAGTGGCATCGGACACCGACCGCCAGGACGCGCTTCAGGTGAACCGCCCTGG 910
Qy 976 AGATGTGAATGTACCGGTGTACTGTCTCTATCTGATGTGATTAACCAAGCTCCCGAGTTTAA 1035
Db 911 GGACCTGAGTGTGCGCTGCAGCTGCTCTCATGTGGACTACCAAGCTCCCGAGTTCAA 970
Qy 1036 ATTAGACCCCGCTAGCTCGACTCCTGGGCATCATACCCAGACTCGTCCAGTGTATCAT 1095
Db 971 ACTGATCCCGCTAGCCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCATTGT 1030
Qy 1096 CCAAGCACTGGGCAATATATTAAACACATAAGCTCCAGGACCTCACGAGCGGGAGTT 1155
Db 1031 CCAGCCCTGTGGCAGTATGTGAAGCAACACAGGCTGCAGAGCTCCCATGACAAAGGAATA 1090
Qy 1156 TGTCACTGTGACAAAGTACCTGCGAGGATCTTTTGAGACTCAACGATTAAGTTTTCAGA 1215
Db 1091 CATCAATGGGACAGATATTTCCAGCAGATTTTGTGATTGTCCTCCCGCTGAAGTTTCTGA 1150
Qy 1216 GATCCTCAGCGGCTCCATGCTTGTGCTTATGTCACCAAGAACTATCATCATTAATCATGT 1275
Db 1151 GATTTCCCGAGCGCTTCACAGCGCTGCTATTTGCCCTTGACCCCAATTGTCATCAACCATGT 1210
Qy 1276 CATCAGTCTTGACCGCAATGATCAGAAAAGACAG-CTTGTTATGACATTGATGTTTCAA 1333
Db 1211 CATCAGCTGGACCCCTTCAGACCCAGAAAGACGCTGTGCTATGACATTGACGTGAG 1270
Qy 1334 GTGGATGACACCTTGAA 1350
Db 1271 GTGGAGGAGCCCATTA 1287

RESULT 7
AAH98319
ID AAH98319 standard; cDNA, 3272 BP.
AC AAH98319;
XX
DT 12-OCT-2001 (first entry)
XX Human EST-derived coding sequence SEQ ID NO: 176.
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX Homo sapiens.
XX WO200154477-A2.
XX 02-AUG-2001.
XX 25-JAN-2001; 2001WO-US02687.
XX 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werfman I;
XX WPI; 2001-476164/51.
DR P-PSDB; AAM23660.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX Claim 1; Page 298-299; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX Sequence 3272 BP; 682 A; 1013 C; 1001 G; 576 T; 0 other;
SQ Query Match 15.0%; Score 425.4; DB 22; Length 3272;
Best Local Similarity 72.6%; Pred. No. 1.6e-106;
Matches 579; Conservative 0; Mismatches 211; Indels 7; Gaps 2;
Qy 556 GATTCTGTGAATCGGTACCAAGATCCAGGCTTATATGATCTCTTGGCTTTTGAAGGAA 615
Db 496 GATTCTGGAGCTGGTCCCGAGTCCAGGCTTACATGGACCTCTTGGCAATTGAGAGGAA 555
Qy 616 ACTGGACACACTATCATAGGAAACGGCTAGATATCCAGAGGCTTGAACGTCCTCAT 675
Db 556 ACTGGATCAAAACCATCATATGCGGAAGCGGTGGACATCCAGAGGCTCTGAAGAGGCCAT 615
Qy 676 TAAGCAAAAACGGAAGCTGCGAATTTTCTTCTAACACTTTCAATCCGGCTAAGTCAGA 735
Db 616 GAAGCAAAAGCGGAGCTGCGACTCTATATCTCCAACACTTTTAACCTCGGAAGCTGA 675
Qy 736 TGCCGAGGATGGGAAGGAGCGGTGGCTTCTGGAGCTTCGGGTAGAAGACCGCTCCT 795
Db 676 TGCTGAGGATTCGACGCGCAGCATTTGCTCTCTGGAGCTACGGGTGGAGGAAGCTCCT 735
Qy 796 GGAGGATTCAGCCTTGTCTCAAATATGATGCCATTAACAAAAGAGGAAGTTCTCTTCTT 855
Db 736 GGATGAT-----GTACGTCCCGGCCCGCCAGCAGCAACAGAAAGCGGAAGTTCTCTCTTT 790
Qy 856 CTTTAAGTCTTGGTGTATTGAACTGGACAAAGACCTGTATGGGCGACACAAACCATCTGGT 915
Db 791 CTTCAAGAGTTTGGTCTATCGAGCTGGACAAAGATCTTTATGGCCCTGACAAACCATCTGT 850
Qy 916 AGAATGGCACAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGG 975
Db 851 TGAGTGGCATCGGACACCGACCGCCAGGACGCGCTTCAGGTGAACCGCCCTGG 910
Qy 976 AGATGTGAATGTACGGTGTACTGTCTTACTGTATGTGGATTACAGGCTCCCGAGTTTAA 1035
Db 911 GGACCTGAGTGTGCGCTGCGCTGCTCTCATGTGGACTCAACGCTCCCGAGTTCAA 970
Qy 1036 ATTAGACCCCGCTAGCTCGACTCCTGGGCTCCATACCCAGACTCGTCCAGTGTATCAT 1095
Db 971 ACTGGATCCCGCTAGCCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCATTGT 1030
Qy 1096 CCAAGCACTGTGGCAATATATTAAAGACACATAAGCTCCAGGACCTCACGAGCGGGAGTT 1155
Db 1031 CCAGGCTCTGTGGCAGTATGTGAAGACCAACAGGCTGCAGGACTCCCATGACAAAGGAATA 1090
Qy 1156 TGTCACTGTGACAAAGTACCTGCGAGGATCTTTTGAGACTCAACGATTAAGTTTTCAGA 1215
Db 1091 CATCAATGGGACAAAGTATTTCCAGCAGATTTTGTGATTGTCCTCCCGCTGAAGTTTCTGA 1150

QY	1216	GATCCCTAGCGGCTCATGCTTGGCTTATGCGACCAACCTATCATTAATCATGCT	1275
Db	1151	GATTTCCCAAGCCCTTACAGCCCTTCTATTGCCCCCTAACCAATTGATCAACCATGT	1210
QY	1276	CATCAGTGTGACCCCAATGATCAGAAAAAGACAG--CTTGTTATGACATTGATGTGAA	1333
Db	1211	CATCAGCGTGGACCTCTCAGACCCAGAAAGAGAGCGTGTCTATGACATTGAGCTGAG	1270
QY	1334	GTGGATGACACCTTGAA	1350
Db	1271	GTGGAGAGACCCATTAA	1287
RESULT 8			
ABL02583			
ID	ABL02583	standard; cDNA; 2010 BP.	
AC	ABL02583;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster	expressed polynucleotide SEQ ID NO 2231.	
XX			
KM	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ss.		
XX			
OS	Drosophila melanogaster.		
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US09231.		
XX			
PR	23-MAR-2000; 2000US-191637P.		
XX	11-JUL-2000; 2000US-0614150.		
XX			
PA	(PEKE) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB58480.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -		
PT			
XX			
PS	Claim 1; SEQ ID NO 2231; 21bp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB101840-AB16175) and the encoded proteins (AB57737-AB872072).		
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
CC			
CC			
XX			
SEQ	Sequence 2010 BP; 509 A; 550 C; 545 G; 406 T; 0 other;		
QY	Query Match	14.6%; Score 414.2; DB 23; Length 2010; Best Local Similarity 68.5%; Pred. No. 1.5e-103; Matches 620; Conservative 0; Mismatches 273; Indels 12; Gaps 3;	
Db	517 AAGAGAAAGAGAGCTGCTGACAAATCTACCTGAAGGATTCGTGAACTGGATCCAGCA 576 580 AAGAGAAAGAGAGCTGCTGACAAATCTGCTGAGAGGTCGGATTCGTGCTCCGCA 639		

OY	577	ATCCAGGCCATATGATCTCTGGCTTTTGAAGAAAGAACTGGACCAAGCTATCAAG	636
Db	640	ATCCAGCGCACTGGATCTCTGACGTTTGGAGCCAAACTGGATCCACATTAATGCC	699
OY	637	GAAGCGCTAGATATCCAAAGCGCTTTGAAACGTCCCATTAAGCAAAAAGAAAGTGGC	696
Db	700	CAAACTGTGGACATCCAGAGGCGCTCAAGCGCCCATGAAGAGAAAGCAAGCTGCG	759
OY	697	AATTTTCATTTCTTAACCTTTCAATCCGGCTAAGTCAGATGCCGAGATGG---GGAA	753
Db	760	CATTTTCATCTCGAACATTTCTATCCAGCAAGAGCCCAAAATGATGCGAGAGGG	819
OY	754	GACGATGAGCTTCCGGGAGCTGGGATGAGAAGAGCGGCTCGTAGAGATTCAAGCTTGTG	813
Db	820	CGCGCTTGCTCTTTGGGAATTTGGGGTGGAGGGTGGCTTTTAGAGAT-----GGCAA	873
OY	814	CAAAATGATGCGCACTAAACAAAAGAGAAATTCCTTCCTTAAAGTCCCTGGTAT	873
Db	874	GGGGGATCCCAATCAAAAGATCAAAAGCAAGTTTGTGCTTTTCAAGTGGCTGGTAT	933
OY	874	TGAATGGAACAAAGACTGTATGGGCCAGAACACATCTGTGATATGSCACAGAGCCG	933
Db	934	CGACCTGGATTAAGAAATGTAACGGATCCGGAACACATCTGATGAGTGGATGCACTCA	993
OY	934	CACTAACCCAGAGACCGATGGCTTCCAGGTAAGACGGCCAGAGATCTGAATATACGGT	993
Db	994	CATCCACCCAGAGACGAGCGGTTCCAGGTGAAGCGGCGGGGATGCAATATGGCTG	1053
OY	994	TACTGTCTTATGATGTGTGATTACAGACCTCCCAAGTTTAAATTAGACCCCGGCTAGC	1053
Db	1054	CACCATCTCCTGTGCTTTGACTACAGCGCGTGCATTTCAAGCTGGAACCCGGACCTGC	1113
OY	1054	TCGACTCTGGGCATCCATACCCAGACTGTCAGATGATCATCCAGCACTGTGGCAATA	1113
Db	1114	CAGGTGCTGGGCGTACACACACAGACCAAGCGCGTATCATATCCGCTGTGGCAGTA	1173
OY	1114	TATTTAAGACATTAAGTCTCAGAGCCCTCAAGAGCGGGAATTTGTCAATCTGTACAAGTA	1173
Db	1174	CATTAAGACGACAAAGCTTACAGAGTGGCCAAGCGGGAATACATCAATTGGCGCAAGTA	1233
OY	1174	CCTGCACAGATCTTTGAGACTCAACGTATGAATTTTCAGAGATCCCTCAGCGGCTCCA	1233
Db	1234	TCTTGAGCAAGTATATTCAAGTTGCGACGCGGATGAATTTGCCGAGTACCGAGGCTTCAA	1293
OY	1234	TGCTCTTATGCGACACAGAACTTATCATATTATCATGTCACTACGTTGACCCGAA	1293
Db	1294	TCCGCTCTGATCCGCGCATCCGATCCGATTGATCAATCATTTTAT---TGAGAGCGTGC	1350
OY	1294	TGATCAGAAAAAGACAGCTTGTATGACATTTGATTTGAAGTGATGACACCTTGAAGAC	1353
Db	1351	AGAAAGAACAGAGACTCCTGTCTACGCAATCTGATGTGAGAGTGGATGATCGCTCAAAA	1410
OY	1354	CCAGATGAATTTCTTTTGTGCTGTCATCTGCAGCGCAACAGAGATTGCTACTAGCAAA	1413
Db	1411	CCAGATGAACAGCTTCTTGATGAGCACTGCGAGCAACAGAGATCCAGGGGCTGGACAC	1470
OY	1414	CAAGA 1418	
Db	1471	CAAGA 1475	
RESULT 9			
ABL02582			
ID ABL02582 standard; cDNA; 4138 BP.			
XX ABL02582;			
XX AC			
XX DT 26-MAR-2002 (first entry)			
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2228.			
XX KW Drosophila; developmental biology; cell signalling; insecticide;			
KW pharmaceutical; gene; ss.			

PI Daifo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tusson O, Yap PE, Amesley SR;
PI Daubertley SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX MPI; 2003-075543/07.
DR P-PSDB; ABP75489.
XX
XX New human secretory proteins and polynucleotides, useful for
PT diagnosing, treating or preventing autoimmune/inflammatory disorders
PT (e.g. AIDS), neurological disorders (e.g. Alzheimer's), or cell
PT proliferations or cancers
PS Claim 1; SEQ ID NO 100; 458bp + Sequence Listing; English.
XX
XX The invention relates to a secretory polynucleotide (designated spm)
CC comprising any of 567 polynucleotide sequences (AB235837-AB236403), a
CC naturally occurring polynucleotide sequence at least 90 % identical to
CC the polynucleotide sequence, a polynucleotide complementary to them or an
CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
CC treating, preventing or diagnosing a disease or condition associated with
CC the expression of functional SPM. These are particularly useful for
CC diagnosing, treating or preventing autoimmune/inflammatory disorders
CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
CC breast, cervix or prostate).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from Wipo at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6809 BP; 1248 A; 1988 C; 2042 G; 1530 T; 1 other;

Query Match 14.4%; Score 408.2; DB 25; Length 6809;
Best Local Similarity 69.7%; Pred. No. 1.3e-101;
Matches 617; Conservative 0; Mismatches 223; Indels 45; Gaps 3;
QY 325 GAGACGAGTATGTTCCAGGAGCCGATGACACTGACCTTCATGGGACCC 384
DB 6712 GGGCCCGGAGATGCTCTGAGACCCGATGCCCAAGGGGCCCCATGGCCCC 6653
QY 385 TGGC-----TATGGGGGGAACCTTCACTCCGACCTGGCCAGTCAAGATGA 438
DB 6652 GGGCTCCCGTCAATGGGACGCCGCCGCTGGACCCCGCCGGCATGGA 6593
QY 439 TCAGTCCGCAAGACCTGCCCCCTCAG-----CAGATCCAGCAGGT 480
DB 6592 GCCCGCCCGCAAGCGAGCGCCCCCGCGGCGAGCCAGCCACAAGCGCAGGCGCA 6533
QY 481 CCAGCGAGCGGGGTCCAAATCGAATCGACAATGCAAGAAAGAAAGATGGCTGACAA 540
DB 6532 GCCGTGCCCAACGCCCGCGCGAGCCGCACTGCGAAGGAGGAGATGGCTGACAA 6473
QY 541 AATTCTACTCAAGAGATTGTAATCTGTATCCAGATCCAGAGCTTATATGATCTTT 600
DB 6472 AATCTTCCCTCAAGAGATTGAGAGCTGTGCTCCCGATGCCAGCTTAATGACCTTT 6413
QY 601 GGGTTTGAAGGAACTGAGACCACTATCATGAGAAAGCGCTAGATATCCAGAGGC 660
DB 6412 GGCATTGAGAGGAACTGATCAAAACATCATGCGGAAGCGGTGAGATCCAGAGAGC 6353
QY 661 CTGTAAGCTCCATTAAAGCAAAAGAGAGCTGGAATTTTCAATTTCAACCTTCA 720
DB 6352 TCTGAAGAGCCCTATAGCAAAAGCGAAGCTGCGACTATATCTTCAACCTTTTAA 6293
QY 721 TCCGCTAAGTCAATGCGAGAGTGGGAGGAGGAGCGTGTGCTTCTGAGAGCTTGCGGT 780
DB 6292 CCTGGGAAGCTGATGCTGAGAGATTCGACGCGCAGCATTTGCTCTGAGAGCTAGCGGT 6233

QY 781 AGAAGAGCGCTCTGAGAGATTACGCTTGCCTTGCCTTGAATATGATGCGCACTAAACAAAGAG 840
DB 6232 GAGAGGGAAGCTCTCTG-----ATGATCCACAGAAACAGAAAGG 6194
QY 841 GAAGTTCTTCTCTTCTTAAAGTCTGTGATGTAAGTGAACAGCAAGACCTGTATGGCC 900
DB 6193 GAAGTTCTTCTTCTTCTTCAAGAGATTGTGATGAGCTGAGCAAAAGATCTTATGGCCC 6134
QY 901 AGACAACCATCTGGTAGAATGCGACAGACCGCCACTACCCAGAGACCGATGGCTTCCA 960
DB 6133 TGACAACCACTCGTTGAGTGGCATCCGACACCCAGACCCAGAGAGAGCGCTTCCA 6074
QY 961 GGTGAAGCGCCAGAGATGTGAATGTACGGTGTCTGCTCCACTGATGCTGATTACCA 1020
DB 6073 GGTGAAGCGCTCTGGGAGCTGAGTGTGCTGACGCTGCTCTCTCATGCTGACTACCA 6014
QY 1021 GCTTCCCACTTAAATTAAGACCCCGCTAGCTGACTCTGGCATTCATACCCAGAC 1080
DB 6013 GCTTCCCACTTCAAACTGAGATCCCGCTAGCCCGGCTGGGGCTGCACACAGAG 5954
QY 1081 TGTCTCAGTATCATCCAGACCTGTGCAATATTTAAGACACTAAGCTTCAGAGACC 1140
DB 5953 CCGCTCAGCATTTGTCAGGCGCTGTGGCAGATGTGAAGACCAACAGGCTGCAGAGCTC 5894
QY 1141 TCAGCAGCGGAGTTGTCTGTCTGTGACAGTACCTGACAGAT 1185
DB 5893 CCAATCAAGGAAATCATCATGAGGAGCAAGATTTCAGCAGGT 5849
RESULT 11
AB054477
ID AB054477 standard; cDNA; 1970 BP.
XX
XX AB054477;
AC
XX
XX 22-AUG-2002 (first entry)
DT
XX
XX
DE Human ovarian antigen HSDY108 cDNA, SEQ ID NO:357.
XX
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200200677-A1.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 07-JUN-2001; 2001WO-US18569.
PF
XX
XX 07-JUN-2000; 2000US-209467P.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Birse CE, Rosen CA;
PI
XX
XX MPI: 2002-147878/19.
DR P-PSDB; ABP41400.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX

PS	Claim 1; SEQ ID No 357; 2922pp; English.	
XX	The invention relates to 2175 novel human ovarian antigens (ABP41054-	
CC	ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also	
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical	
CC	to the sequences of the invention. The invention additionally relates to	
CC	recombinant vectors and host cells comprising human ovarian antigen	
CC	polynucleotides, antibodies against human ovarian antigens, and the use	
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,	
CC	treating, prognosing or preventing various ovary and/or breast-related	
CC	disorders. Such conditions include ovarian cancer and breast cancer, and	
CC	metastatic tumours of ovarian or breast origin, reproductive system	
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,	
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine	
CC	disorders, infectious conditions (e.g., chlamydia, HIV, toxoplasmosis, and toxic	
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and	
CC	vaginitis), immune disorders (e.g., congenital and acquired	
CC	immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),	
CC	blood-related disorders (e.g., anaemia), cardiovascular disorders,	
CC	respiratory disorders, neurological disorders, gastrointestinal disorders	
CC	and urinary system disorders. Ovarian antigen polypeptides and	
CC	polynucleotides may also be used in screening for compounds which	
CC	modulate ovarian antigen expression or activity. The polynucleotides may	
CC	further be used for gene therapy, chromosome mapping, in the	
CC	identification of individuals and in forensic analysis, and the	
CC	polypeptides may be used as food additives or to prepare antibodies	
CC	useful in disease diagnosis, drug targeting and phenotyping. The present	
CC	sequence represents cDNA encoding a human ovarian antigen of the	
CC	invention.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 1970 BP; 463 A; 590 C; 522 G; 389 T; 6 other;	
	Query Match 14.0%; Score 397; DB 24; Length 1970;	
	Best Local Similarity 69.3%; Pred. No. 8.5e-99;	
	Matches 615; Conservative 3; Mismatches 223; Indels 46; Gaps 4;	
QY	323 CCGACACAGGTATCTTCCAGGAGCCGATGACACCTCAGGACCTTCCATGGGACCC 382	
DB	172 CAGCCCGCGGATCCGCTCTGGAGCCGCGATGCCCCACAGGGGGCCCATGGGCCCC 231	
QY	383 CTGTGC-----TATGGGGGAAACCTTTCAGTCCACCTGGCTGGCCAGTCAGGGATG 436	
DB	232 CCGGCTCTCCCGTACATGGGAGCCCGCCGCTGCGACCCGCGCTGGCCCGGGCATG 291	
QY	437 GATCAGTCCCGCAGAGACCTGCCCCCTCAG-----CAGATCCAGCAG 478	
DB	292 GAGCCCGCCGCAAGCAGAGCAGCGCCCCCGCCGGGAGCCAGGCGACAGAGCCAGGGC 351	
QY	479 GTCCAGCAGCAGCGCGTCCAAATCGAAACCAATGCAAGAAAGAAAGAGAGTGGCTGAC 538	
DB	352 CAGCCGGTGCACCGCCCGCCGCGGAGCGCGAGTGCCTGCAAGAGAGGAAAGTGGCTGAC 411	
QY	539 AAAATTCCTACCTCAAGGATTCGTGAACTGTGTACAGAAATCCAGGCGCTATATGGATCTC 598	
DB	412 AAAATCTCCCTCAAGGATTCGGAGTGTGTCGCCGAGTCCCGAGTCCAGGCTTACATGGACCTC 471	
QY	599 TTGGCTTTTGAAGGAAACTGGACAGACTATCATGAGGAAACGGCTAGATATCCAAGAG 658	
DB	472 TTGGCATTTGAGAGGAAACTGGATCAAAACCATCATGCGGAGCGGCTGGACATCCAGGAG 531	
QY	659 GCGTTGAACAGTCCCATTAAGCAAAAACGGAGCTGGCAATTTTTCATTCTCAACATTTTC 718	
DB	532 GCTCTGAAGAGGCCCATGAAGCAAAAGCGGAAGCTGCGACTATATCTCAACACTTTT 591	
QY	719 AATCCGGCTTAAGTCAGATGCCAGATGGGGAAGGAGCGGTGCTTCTGGGAGCTTCGG 778	
DB	592 AACCTTGGGAG-CTGATGCTGAGGATTCGCGAGCAGCATTTGCCCTCTCTGGGAGCTACGG 650	
QY	779 GTAGAGGACCGGCTCCTGGAGGATTCAGCCCTTGTCCAAATATGATGCGCACTAAACAAAAG 838	

Db

651 GTGAGGGGAAGCTCTCTGG-----ATGATCCAGCAAAACAGAAG 689

QY

839 AGGAAGTCTCTTCTCTTTTAACTCCTTGGTGATTGAACTGGACAAAGACCTGTATGGG 898

Db

590 CGGAAGTCTCTTCTTCTTCAAGAGTTGGTCACTGAGCTGGACAAAGATCTTTATGGC 749

QY

899 CCAGAACCACTCTGGTAGAATGGCACAGGACCCCACTACCCAGGAGACCGATGGCTTC 958

Db

750 CCTCACCAACCACTCTGGTAGAATGGCACAGGACCCCACTACCCAGGAGACCGATGGCTTC 809

QY

959 CAGGTGAAGCGGCAGAGATGTAATGTACGGTGTACTGTCTTACTGATGCTGGATTAC 1018

Db

810 CAGGTGAAGCGGCCTGGGGACCTGAGTGTGCGCTGCACTGCTCTCTCATGCTGGACTAC 869

QY

1019 CAGCCTCCCACTTTAAATAGACCCCGCTAGCTCGACTCCTGGGCATCCATACCCAG 1078

Db

870 CAGCCTCCCACTTTCAAACTGGATCCCGCTAGCCCGCTGCTGGGCTGCCACACAG 929

QY

1079 ACTCGTCCAGTGATCATCCAAAGCACTGTGGCAATATATTAAAGACACATAAGCTCCAGGAC 1138

Db

930 AGCCGCTCAGCCATTGTCAGGCCCTGTGGCAGVATGGAAGACCAACAGGCTGSGAGAC 989

QY

1139 CTTACAGAGCGGGAGTTTGTCTCATCTGTACAGTACCTGCGAGCAT 1185

Db

990 TCCCATGACAAAGGAATACATCAATGGGACAAAGTATTTCCAGCAGGT 1036

RESULT 12

AAZ80712/C

ID AAZ80712 standard; cDNA; 452 BP.

XX

AAZ80712;

DT

07-APR-2000 (first entry)

XX

Human colon cancer cell line SW480 cDNA clone SEQ ID NO:796.

DE

Human; gene expression product; diagnosis; tumour; colon cancer;

XX

colorectal adenocarcinoma; cell line SW480; cell proliferation;

KW

cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;

KW

hyperplasia; ds.

XX

Homo sapiens.

OS

WO9964576-A2.

PN

16-DEC-1999.

XX

09-JUN-1999; 99WO-IB01062.

PF

10-JUN-1998; 98US-0088801.

XX

(FARB) BAYER CORP.

PR

Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;

PI

Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;

XX

Schlegel R;

DR

WPI; 2000-087220/07.

XX

Novel nucleic acids, used to develop products for the diagnosis and

PT

treatment of disorders involving unwanted cell proliferation,

PT

particularly cancers, especially colon cancer

XX

Claim 15; Page 448; 469pp; English.

PS

AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from

CC

the human colorectal adenocarcinoma (colon cancer) cell line SW480. The

CC

cDNA clones can be used to generate antisense oligonucleotides which

CC

can be used for antisense therapy. Methods and products from the present

CC

invention can be used for identifying and/or classifying cancerous cells

CC

present in a human tumour, particularly in solid tumours, e.g.

CC

carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones

CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.

XX Sequence 452 BP; 99 A; 124 C; 104 G; 122 T; 3 other;

Query Match 13.7%; Score 388.8; DB 21; Length 452;
Best Local Similarity 97.9%; Pred. No. 7.2e-97;
Matches 415; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 566 CTGTGACCAAGATCCAGGCTTATGATCTCTGGCTTTGAAAGAACTGACCG 625
DB 423 CAGGATACCAAGATCCAGGCTTATGATCTCTGGCTTTGAAAG-ACCTGACCG 365
QY 626 ACTATCATGAGAAACGGCTAGATATCCAGAGGCGCTTGAACGCTCCCTTAAGCAAAA 685
DB 364 ACTATCATGAGG-AAACGGCTAGATATCCAGAGGCGCTTGAACGCTCCCTTAAGCAAAA 306
QY 686 CCGAAGCTGCGAATTTCTATTCTAACACTTTCAATCCGCTAAGTCAGATGCCAGAGAT 745
DB 305 CCGAAGCTGCGAATTTCTATTCTAACACTTTCAATCCGCTAAGTCAGATGCCAGAGAT 246
QY 746 GGGGAAAGGAGCGGTGCTTCTGGAGCTTCGGGTGAAAGACGCGCTCTGAGGATTCA 805
DB 245 GGGGAAAGGAGCGGTGCTTCTGGAGCTTCGGGTGAAAGACGCGCTCTGAGGATTCA 186
QY 806 GCTTGTCCAAATATGATGCCACTTAACAAAGAGAAATCTCTCTCTTTTAAGTCC 865
DB 185 GCTTGTCCAAATATGATGCCACTTAACAAAGAGAAATCTCTCTCTTTTAAGTCC 126
QY 866 TTGATGATGAACTGACCAAGACCTGTATGGGCCAGACCAACATCTGTAGATGGCAC 925
DB 125 TTGATGATGAACTGACCAAGACCTGTATGGGCCAGACCAACATCTGTAGATGGCAC 66
QY 926 AGGACCGCCACTTACCCAGAGACCGATGGCTTCCAGGTGAAGCGCCAGAGATGTGAAT 985
DB 65 AGGACCGCCACTTACCCAGAGACCGATGGCTTCCAGGTGAAGCGCCAGAGATGTGAAT 6
QY 986 GTAC 989
DB 5 GTAC 2

RESULT 13
AAZ80561/c
ID AAZ80561 standard; cDNA; 690 BP.
XX
AC AAZ80561;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:645.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Asle JH, Burgess CC, Bushnell SE;
PI Carroll E, Carino TJ, Derci A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;

XX WPI; 2000-087220/07.
DR
XX

PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer

PS Claim 15; Page 387; 469pp; English.

CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g.
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.

XX Sequence 690 BP; 165 A; 168 C; 170 G; 169 T; 18 other;

Query Match 10.9%; Score 310.2; DB 21; Length 690;
Best Local Similarity 99.0%; Pred. No. 4.7e-75;
Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1409 GACAAACAGACATACATGATGTGGTAAACCCAGAGAGAGACCGCGAGCTGATTC 1468
DB 315 GACCTCAAGACATACATGATGTGGTAAACCCAGAGAGAGACCGCGAGCTGATTC 256
QY 1469 TACTTCAGACCCCTGGGCTCAGAGAGCTGTGCGCATCTTCTACTCCAGGTGACGAG 1528
DB 255 TACTTCAGACCCCTGGGCTCAGAGAGCTGTGCGCATCTTCTACTCCAGGTGACGAG 196
QY 1529 AGACGACAGAAATTAGACCAAGCCCTGGGAATCCGGAATACATAGAGCCCTCTCCACAGC 1588
DB 195 AGACGACAGAAATTAGACCAAGCCCTGGGAATCCGGAATACATAGAGCCCTCTCCACAGC 136
QY 1589 CCTGATTTGACATGACCAATTTCTTATTTGGGCGCTGTGCTGCTCCATAGATCTG 1648
DB 135 CCTGATTTGACATGACCAATTTCTTATTTGGGCGCTGTGCTGCTCCATAGATCTG 76
QY 1649 CCTGATTTGACATGACCAATTTCTTATTTGGGCGCTGTGCTGCTCCATAGATCTG 1708
DB 75 CCTGATTTGACATGACCAATTTCTTATTTGGGCGCTGTGCTGCTCCATAGATCTG 16
QY 1709 AAGAGGCTCTCAAA 1723
DB 15 AAGAGGCTCTCAAA 1

RESULT 14

ABX34564
ID ABX34564 standard; cDNA; 2581 BP.

XX
AC ABX34564;
XX

DT 13-FEB-2003 (first entry)
XX

DE Human mddt cDNA SEQ ID 125.
XX

KW MDDT; human; disease detection and treatment molecule polypeptide;
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianemic; antipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis; gene; ss.

XX Homo sapiens.
XX
OS
XX
PN WO200279449-A2.

XX	10-OCT-2002.	QV	879	TGACAAAGACCTGTATGGGCCAGACCAACCATCTGGTAG-----	917
PD					
XX		Db	884	TGGACAAGGAGCTGTACGGGCTGACAATCACCTGGTGGAGCCATGCTTCCTCCCTGCC	943
PF	27-MAR-2002; 2002WO-US09944.				
XX					
XX	28-MAR-2001; 2001US-279619P.	QV	918	AATGGCAAGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAG	977
PR	29-MAR-2001; 2001US-280067P.				
PR	29-MAR-2001; 2001US-280068P.	Db	944	AGTGGCACCGGATGCCCAACCAACCCAGGAGACAGATGGCTTCCAAGTAAAAACGGCTGGAG	1003
PR	16-MAY-2001; 2001US-291280P.				
PR	17-MAY-2001; 2001US-291829P.	QV	978	ATGTGAATGTACGGTGTACTGCTCTACTGTCTGTGGATTACACAGCCTCCAGTTTAAAT	1037
PR	17-MAY-2001; 2001US-291849P.				
PR	19-JUN-2001; 2001US-299428P.	Db	1004	ACCTCAACGTCAGTGACCCCTCTCTCTCATCTGGATCATCAGCCTCCCGAGTACAAAT	1063
PR	20-JUN-2001; 2001US-299776P.	QV	1038	TAGACCCCGCTAGCTCGACTCTCTGGGCATCCATACCCAGACTCGTCCAGTGTATCC	1097
PR	20-JUN-2001; 2001US-300001P.				
XX	(INCY-) INCYTE GENOMICS INC.	Db	1064	TGGACCCCGATTGGCAAGGCTGCTGGAGTGCACACGACGAGGCGCCCATCATGC	1123
PA					
XX	Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;	QV	1098	AAGCAGTGTGGCAATATTAAAGACA CATAGCTCCAGGACCCCTCAGAGCGGAGTTTG	1157
PI	Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;				
PI	Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;	Db	1124	AGGCCCTGTGGCTTTACATCAAGCAACACAGCTGCAGGATGGCGACGAGCGGAGTACA	1183
PI	Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;				
PI	Flores V, Marwana R, Lo A, Lan RY, Urashka ME;	QV	1158	TCATCTGTGACAAGTACTGCAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGA	1217
XX					
XX	WPI; 2003-058431/05.	Db	1184	TCAACTGCAACCGTTACTTCCGCCAGATCTTCAGTTGTGGCCGACTCCGTTCTCCGAGA	1243
DR	P-PSDB; ABU11574.				
XX		QV	1218	TCCCTCAGCGCTCCATGCTTGTCTTATGCCACAGAACCTTATCATCATTAATCATGTCA	1277
XX					
XX	New purified disease detection and treatment molecule proteins and	Db	1244	TTCCCATGAAGCTGGCAGGTTGCTGCAGCATCCAGACCCCATTTGTCTCATCAACCATGTCA	1303
PT	polynucleotides, useful for diagnosing, treating or preventing cancers	QV	1278	TCAGTGTGACCCGAATGATCAGAAAAGACAGCTTGTGTATGATGATTTGAAGTGG	1337
PT	(e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis				
PT	or hepatitis	Db	1304	TTAGTGTGACCCCTTAACGACCAAGAAAGACAGCTGTTCAGCATCATGATGTGGAGGTGG	1363
XX					
PS	Claim 1; SEQ ID NO 125; 339pp + Sequence Listing; English.	QV	1338	ATGACACTTGAACACCCAGATGAATCTTTTCTGCTGTCCACTGCCAGCAACAGGAGA	1397
XX					
CC	This invention describes a novel disease detection and treatment molecule	Db	1364	ACGACCCACTGAAGGCCCAATAGCAATTTTCTGGCTCTACCAACCAATCAGCAGGAGA	1423
CC	polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,	QV	1398	TTGCTACTCTAGACAACAAGA 1418	
CC	osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,				
CC	antianemic, antipsoriatic and hepatotropic activity. The polynucleotides	Db	1424	TCGCTCCCTTGATGTCAAGA 1444	
CC	and the polypeptides of the invention can be used for gene therapy,				
CC	protein replacement therapy and are useful for treating a variety of				
CC	diseases or conditions. These polypeptides or polynucleotides are				
CC	particularly useful for diagnosing, treating or preventing cell				
CC	proliferative disorders (e.g. cancers including adenocarcinoma,				
CC	leukemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's				
CC	disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's				
CC	syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or				
CC	hepatitis. ABX3440-ABX3485 encode the MDDT polypeptides represented in				
CC	ABU11450-ABU11845, described in the disclosure of the invention.				
CC	NOTE: The sequence data for this patent did not form part of the printed				
CC	specification, but was obtained in electronic format from WIPO at				
CC	ftp.wipo.int/pub/published_pct_sequences.				
XX					
SQ	Sequence 2581 BP; 572 A; 652 C; 737 G; 620 T; 0 other;				
	Query Match 10.5%; Score 298.6; DB 25; Length 2581;				
	Best Local Similarity 64.2%; Pred. No. 1.5e-71;				
	Matches 514; Conservative 0; Mismatches 254; Indels 33; Gaps 3;				
QV	651 TCCAGAGCCCTTGAACGTCCATTAGCAAAAACGGAAGCTGCGAATTTTCATTCTA	710			
Db	644 TCCAGTTGTGATGCTTCTCTCTTTAGCAAAAGCGAAAGCTTCGGATCTACATTTCCA	703			
QV	711 ACACCTTTCATCCGCTAAGTCAGATCGCGAGGATGGGGAAGGACGGTTCCTCGG	770			
Db	704 ATACCTTCAGTCCGAGACGGCGGAGGGATAGTGCAGGAACCTCAGGACCCCTGGG	763			
QV	771 AGCTTCGGGTAGAAG- ---GACGGCTCCTGGAGGATTCAGCCTTGTCCAAA-----T	818			
Db	764 GAAACCCAGCAGGGGACAAAGTGGCTTCTTGGAACTCCGAGTGAAGGAAAACCTGTGG	823			
QV	819 ATGATGCCACTAAACAAAAGAGGAAGTCTCTCTCTCTTTTAAGTCCTTGGTATTGAAC	878			
Db	824 ATGATCCTAGCAACAGAGGAAGGAGTTTCTTCTTCTTCTTTTAAGAGCCTCGTCATTTGAC	883			

Human BRG1-associated factor (BAF) 60c3 DNA.

Human; nuclear receptor cofactor; proliferator-activated receptor; RXR; PPAR-gamma; retinoid X receptor; cell proliferative disorder; infection; metabolic disorder; cardiovascular disorder; inflammatory disease; acne; neurodegenerative disease; Parkinson's disease; psoriasis; cancer; HIV; renal disease; atherosclerosis; Alzheimer's disease; diabetes; vaccine; osteoporosis; human immunodeficiency virus; BRG1-associated factor; BAF; gene therapy; ds.

Homo sapiens.

Key Location/Qualifiers
1..753
CDS/product= "BAF60c3 protein"
/transl_except= (pos:598..609, aa:Arg-Gln)
/transl_except= (pos:610..618, aa:Met-Pro)
/note= "No stop codon"
/partial

WO200294877-A2.

28-NOV-2002.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 12:26:42 ; Search time 138 Seconds
(without alignments)
9086.735 Million cell updates/sec

Title: U66617
Perfect score: 2841
Sequence: 1 GAATTCGGCTATCCCATG.....TGGGCAAAAGGGGATTC 2841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	388.8	13.7	452	3	US-09-328-111-796 Sequence 796, App
C 2	310.2	10.9	690	3	US-09-328-111-645 Sequence 645, App
3	76	2.7	450	4	US-09-397-787-279 Sequence 279, App
4	69.6	2.4	283	4	US-09-313-294-3534 Sequence 3534, App
5	66.8	2.4	7218	1	US-08-232-463-14 Sequence 14, Appl
C 6	51	1.8	441529	3	US-09-103-840A-1 Sequence 1, Appl
C 7	50.4	1.8	152331	3	US-09-128-155-16 Sequence 16, Appl
C 8	50.2	1.8	4403765	3	US-09-103-840A-2 Patent No. 5244792
9	49.8	1.8	3472	6	US-09-103-840A-2 Sequence 2, Appl
10	49.4	1.7	4403765	3	US-09-103-840A-1 Sequence 1, Appl
11	49.4	1.7	441529	3	US-09-103-840A-1 Sequence 1, Appl
C 12	49	1.7	2598	4	US-09-816-094-1 Sequence 3, Appl
C 13	49	1.7	7301	4	US-09-816-094-1 Sequence 3, Appl
C 14	46.2	1.6	30001	1	US-08-125-468-1 Sequence 1, Appl
C 15	46.2	1.6	30001	2	US-08-474-933-1 Sequence 1, Appl
C 16	45.6	1.6	747	4	US-09-252-991A-3590 Sequence 3590, App
C 17	45.6	1.6	1191	4	US-09-252-991A-3692 Sequence 3692, App
C 18	45.6	1.6	1845	4	US-09-252-991A-3639 Sequence 3639, App
19	45	1.6	319	3	US-09-165-264-8 Sequence 8, Appl
20	44.6	1.6	6453	3	US-08-306-691B-14 Sequence 14, Appl
21	44.6	1.6	6453	3	US-09-209-668-10 Sequence 10, Appl
22	44.6	1.6	6453	3	US-09-356-952-8 Sequence 8, Appl
23	44.4	1.6	6816	4	US-09-404-650-1 Sequence 3, Appl
24	44.4	1.6	6855	4	US-09-404-650-3 Sequence 13, Appl
25	44.2	1.5	320	3	US-09-165-264-13 Sequence 14, Appl
26	43.8	1.5	320	3	US-09-165-264-14 Sequence 5, Appl
27	43.8	1.5	3772	4	US-09-962-665-5

C 28	43.6	1.5	3047	4	US-09-016-434-1465 Sequence 1465, App
29	43.2	1.5	983	4	US-09-512-363-1 Sequence 1, Appl
30	43.2	1.5	983	4	US-09-176-200-1 Sequence 1, Appl
31	43	1.5	318	3	US-09-165-264-12 Sequence 12, Appl
32	43	1.5	8147	4	US-09-514-247A-9 Sequence 9, Appl
33	42.6	1.5	320	3	US-09-165-264-7 Sequence 2, Appl
34	42.6	1.5	1611	3	US-08-909-742-2 Sequence 2, Appl
35	42.6	1.5	1611	3	US-09-412-289-2 Sequence 2, Appl
C 36	42	1.5	400	4	US-09-056-556-179 Sequence 179, App
C 37	42	1.5	400	4	US-09-072-596-174 Sequence 174, App
C 38	42	1.5	962	4	US-08-765-907A-16 Sequence 16, Appl
C 39	42	1.5	1052	2	US-08-403-852D-10 Sequence 10, Appl
C 40	42	1.5	1052	3	US-08-510-646B-10 Sequence 10, Appl
C 41	42	1.5	1052	3	US-09-231-818-10 Sequence 10, Appl
C 42	42	1.5	2174	4	US-09-613-444-1 Sequence 1, Appl
C 43	42	1.5	2888	4	US-08-765-907A-1 Sequence 1, Appl
44	41.6	1.5	985	4	US-09-056-556-182 Sequence 182, App
45	41.6	1.5	985	4	US-09-072-596-177 Sequence 177, App

ALIGNMENTS

```
RESULT 1
US-09-328-111-796/c
Sequence 796, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Aetle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Carliho, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 796
LENGTH: 452
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(452)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-796

Query Match      13.7% Score 388.8; DB 3; Length 452;
Best Local Similarity 97.9% Pred. No. 4,1e+100;
Matches 415; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
```

666 CCGAAGCTGCGAATTTTCTTACACTTTCATCCGCTTAAGTCCGAGAT 745

Db 305 CGGAAGCTCGCAATTTTCATTTTCAACACTTTTCAATCCGGCTAAGTCAGATGCCGAGGAT 246
QY 746 GGGGAAGGACGGTGGCTTCTGGGAGCTTCGGGTAGAGGACGGCTCCTGGAGGATTC A 805
Db 245 GGGGAAGGACGGTGGCTTCTGGGAGCTTCGGGTAGAGGACGGCTCCTGGAGGATTC A 186
QY 806 GCCTTGTCCAAATATGATGCCACTAAACAAAGAGAGAAATTCCTCTTCCTTTTAAAGTCC 865
Db 185 GCCTTGTCCAAATATGATGCCACTAAACAAAGAGAGAAATTCCTCTTCCTTTTAAATCC 126
QY 866 TTGGTGATTAAGTGGCAAGAGACCTGTATGGGCCAGACAACCATCTCTGTAGATGGCAC 925
Db 125 TTGGTGATTAAGTGGCAAGAGACCTGTATGGGCCAGACAACCATCTCTGTAGATGGCAC 66
QY 926 AGGACCGCCACTACCCAGGAGACCGATGCCTTCCAGGTGAAGCGCCAGGAGATGTAAT 985
Db 65 AGGACCGCCACTACCCAGGAGACCGATGCCTTCCAGGTGAAGCGCCAGGAGATGTAAT 6
QY 986 GTAC 989
Db 5 GTAC 2

RESULT 2

US-09-328-111-645/c
; Sequence 645, Application US/09328111
; Patent No. 6262333

GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 645
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(690)
; OTHER INFORMATION: n = A,T,C or G

Query Match 10.9%; Score 310.2; DB 3; Length 690;

Best Local Similarity 99.0%; Pred. No. 9e-78;

Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1409 GACAACAGACAAATCACTGATGTGTGGTAAACCCAGAGGAGGACGCCGAGCTGAGTTC 1468

Db 315 GACCTCAAGACAATCACTGATGTGTGGTAAACCCAGAGGAGGACGCCGAGCTGAGTTC 256

QY 1469 TACTTCCAGCCCTGGGCTCAGAGGCTGTGTGCCGATCTTCTACTTCCAAAGTGCAGCAG 1528

Db 255 TACTTCCAGCCCTGGGCTCAGAGGCTGTGTGCCGATCTTCTACTTCCAAAGTGCAGCAG 196

QY 1529 AGACGACAAGAAATTAGACAAGCCCTGGGAATCCGGAATACATAGGCGCTCTCCACAGC 1588

Db 195 AGACGACAAGAAATTAGACAAGCCCTGGGAATCCGGAATACATAGGCGCTCTCCACAGC 136
QY 1589 CCTGATTCGACTGCACCAATTTCTTGAATTTGGGCCCTGTGCTGCTCATAGTATCTG 1648
Db 135 CCTGATTCGACTGCACCAATTTCTTGAATTTGGGCCCTGTGCTGCTCATAGTATCTG 76
QY 1649 CCTTGGTCTTGTCTTGGGGCGTTCCAGGGGATGTCTTGGTTCAAAGGACCAACACAGAATG 1708
Db 75 CCTTGGTCTTGTCTTGGGGCGTTCCAGGGGATGTCTTGGTTCAAAGGACCAACACAGAATG 16
QY 1709 AAGAGGCTCTCACAA 1723
Db 15 AAGAGGCTCTCACAA 1

RESULT 3

US-09-397-787-279
; Sequence 279, Application US/09397787
; Patent No. 6468758

GENERAL INFORMATION:

; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; FILE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 279
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(450)
; OTHER INFORMATION: n = A,T,C or G

US-09-397-787-279

Query Match 2.7%; Score 76; DB 4; Length 450;

Best Local Similarity 61.8%; Pred. No. 1.3e-11;

Matches 168; Conservative 0; Mismatches 80; Indels 24; Gaps 2;

QY 323 CCGAGACGAGTATGTTGCCAGGCGCGAATGACCTCAGGACCTTCCATGGGACCC 382

Db 178 CAGCGCCCGGGATGCGCTCTGGAGCCCGGATGCCACCAGGGCGGCCCATGGGCCCC 237

QY 383 CCTGGC-----TATGGGGGAACCTTTCAGTCCGACCTGGCCTGGGCCAGTCAGGGATG 436

Db 238 CCGGGCTCCCGTACATGGGCGAGCCCGCGCTGCGACCCGCGCTGGCCCGCGGGCATG 297

QY 437 GATCAGTCCCGCAAGACCTGCCCCCTCAG-----CAGATCCAGCAG 478

Db 298 GAGCCCGCCGCAAGCGAGCAGCGCCCCCGCGGCGAGAGCCAGGACACAGCCAGGCC 357

QY 479 GTCCAGCAGCGGGGTCCAAAATCGAAACCAACATCGAAAGAAAAGAGATGGCTGAC 538

Db 358 CAGCCGGTGGCCACCCCGCCCGCGGAGCGCGAGTCCCAAGAGAGGAAGATGGCTGAC 417

QY 539 AAAATTCTTACCTCAAAGGATTCGTGAATGGT 570

Db 418 AAAATCTCTCTCAAAGGATTCGGGAGCTGGT 449

RESULT 4

US-09-313-294A-3534

; Sequence 3534, Application US/09313294A

; Patent No. 6476212

GENERAL INFORMATION:

; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.

APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 3534
LENGTH: 283
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700611809H1
US-09-313-294A-3534

Query Match 2.4%; Score 69.6; DB 4; Length 283;
Best Local Similarity 53.3%; Pred. No. 6.5e-10;
Matches 147; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 1077 AACCTGCTCAGTGCATCCAAAGCAGCTGGCAATATTTAAGACATAGCTCCAGG 1136
DB 5 AACCTGCCACGGGAGTGTCTGCTCCCTCGCAGATATTAAGCAAGAGCTTCAGA 64
QY 1137 ACCCTCAGCAGCGGAGTTGTCTATCTGACAGTACCTGCAGCATCTTGAAGACTC 1196
DB 65 ACCCAATATCTCTTCTCTTCACTGTGACCCCTCAATTGAAGAGTGTGGGAGG 124
QY 1197 AACGATGAGTTTTCAGAGATCCCTCAGCGGCTCCATCCTTCTTATGCAACGAGAC 1256
DB 125 ACAAGCTCAAGTTTGCAATGCTGTCAAGAGATATCTCAGCATCTGTCTGCCGCCAC 184
QY 1257 CATATCATTTATATATGTCATCAGTGTGACCCGAATATCAGAAAAAGACAGCTTT 1316
DB 185 CCATCAATTTGAGCACAAGATTAACTGTGCGGAGATGAACGCAATGACAGTGTCT 244
QY 1317 ATGACATTGATGTGAGTGAATGATGACACCTTGAGA 1352
DB 245 ATGATGTCTGTGTGATGTCTTCCCTTCCCGCTGAGA 280

RESULT 5
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZspc-Fls
US-08-232-463-14

Query Match 2.4%; Score 66.8; DB 1; Length 7218;
Best Local Similarity 7.8%; Pred. No. 2.3e-08;
Matches 35; Conservative 232; Mismatches 179; Indels 0; Gaps 0;

QY 1899 TGGTCTTCAAGGCGACCGCTACTGATCTTACCTTCAGAGACCCAGAGTGGCTTC 1958
DB 1004 TCGTTGCCATACGCTCAGAGATTATCCGAGCTTGCTCAGTGCAGGAGCTTCC 1063
QY 1959 GCTCCTTCCAAAGCTCAGGCGCTGGGCACTATTAAGCTTGAATCTTGCGCTCC 2018
DB 1064 GATT 1123
QY 2019 TGATACAGAAATCCAAATTCCTTCTCCCTCCACAGGTTTGGAAACAACTCCCTTCA 2078
DB 1124 YY 1183
QY 2079 CTGTGCTCCCTGACACTACAGAAACCTGTTCTTGCTCCACTGAGCCCAAGTACG 2138
DB 1184 YY 1243
QY 2139 TCCCGAGCTCGGCTGGGCGCTGTCAGTCTCTCTCACTCTTGTAGTGGGTCAC 2198
DB 1244 YY 1303
QY 2199 ATCAGTATGAGTTTGTCTTTATGCTCCCTCCAGACACTCCCTGCTGCCCTT 2258
DB 1304 YY 1363
QY 2259 TGTGATTCCTCAGATCTGCCCTAATCCCGGCAATTTGGGTGGGGAATCTTGCTTCC 2318
DB 1364 YY 1423
QY 2319 CTTTCAAGCCCGGAGTCTCATCT 2344
DB 1424 YYYYYYYYYGTACCAAAATCTTCT 1449

RESULT 6
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA


```

1 OPERATING SYSTEM: PC-DOS/MS-DOS
2 SOFTWARE: PatentIn Release #1.0, Version #1.25
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/474,933
5 FILING DATE:
6 CLASSIFICATION: 435
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: US 08/125,468
9 FILING DATE: 22-SEP-1993
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Tsevdos, Estelle J
12 REGISTRATION NUMBER: 31,145
13 REFERENCE/DOCKET NUMBER: 31,255-02
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: (201)831-3241
16 TELEFAX: (201)831-3305
17 INFORMATION FOR SEQ ID NO: 1:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 30001 base pairs
20 TYPE: nucleic acid
21 STRANDEDNESS: single
22 TOPOLOGY: linear
23 MOLECULE TYPE: DNA (genomic)
24 US-08-474-933-1

```

Query Match	1.6%	Score 46.2	DB 2	Length 30001

Matches 142; Conservative 0; Mismatches 138; Indels 3; Gaps 1;

QY 19 AGTCTCGCTGCCCTGAGCCTCCCGTGCCCGCCGGCCGGGGAACAGGCGGGCCG 78

Db 23965 AGGCGGCGCTGGAAGGCGCTGGCCAGGCGCCGCGCACGTGCGGAGGCAACCGGC 23906

79 GGGGCGCTCGGGGCGGGGAGTTCGGTTCGGTCTTGTGCGGCTGCATCGCG 138

Dh 23905 TGATGCTGGCCGCCGGTGGACGGCTCGCTTGCCCCCTGTGTCGTCCTGACC 23846

139 GCTCCGGAGATGGCGCCCGGGCGGTTCCAGTCTGTGGCTCCAAGCGGCGCCG 198

Db 23845 GCAGCGGCAAGTACGCCCGGGCGGACCGGAGCGGCTTCTGACGCCG 23786

QY 199 GAGCCTCAGGAGGG--CGGGCGGCTGCTGCTTGGGCCCGCGGAATCCGGGCC 255

Db 23785 ACGCCTCCGGCGGGTGCCCGGCGAGGGCGCGCACTGCTCGTGTGAGACCCCGACG 23726

256 TCCTGTGGAATGGCCCGGCTCCGGCTCAGGGCTGTACCGC 298

Db 23725 GCGCCGCCGAGCGCGGCAGCACCGCCACTACGGGGCGATCGC 23683

Search completed: October 11, 2003, 15:44:33
Job time : 162 secs

Job time : 162 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 14:37:03 ; Search time 525 Seconds
(without alignments)
14041.636 Million cell updates/sec

Title: U66617

Perfect score: 2841
Sequence: 1 GAATTCGCGCTATCCCATAG.....TGGCACAAGGCGATTTC 2841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1457.4	51.3	2017	14 US-10-102-806-137	Sequence 137, App
2	488.8	17.2	606	10 US-09-920-300A-11	Sequence 11, App1
3	488.8	17.2	606	12 US-10-099-926-11	Sequence 11, App1
4	488.8	17.2	606	13 US-10-033-528-11	Sequence 11, App1
5	388.8	13.7	452	10 US-09-879-536-796	Sequence 796, App
6	310.2	10.9	690	10 US-09-879-536-645	Sequence 645, App
7	263.8	9.3	485	11 US-09-918-995-16164	Sequence 16164, A
8	237.8	8.4	474	11 US-09-918-995-26465	Sequence 26465, A
9	222.6	7.8	447	11 US-09-918-995-28419	Sequence 28419, A
10	174.4	6.1	495	11 US-09-918-995-2832	Sequence 2832, App
11	123.4	4.3	310	14 US-10-070-676-23390	Sequence 23, App1
12	113.6	4.0	285	10 US-09-960-352-12390	Sequence 12390, A
13	104	3.7	356	10 US-09-783-590-3701	Sequence 3701, App
14	102.6	3.6	420	11 US-09-918-995-34033	Sequence 34033, A
15	97.4	3.4	462	11 US-09-918-995-28488	Sequence 28488, A
16	91	3.2	440	11 US-09-918-995-27234	Sequence 27234, A

17	91	3.2	471	11 US-09-918-995-29605	Sequence 29605, A
18	82.6	2.9	15500	9 US-09-764-869-2109	Sequence 2109, App
19	82.6	2.9	15500	14 US-10-091-504-2109	Sequence 2109, App
20	79.6	2.8	740	9 US-09-770-149-607	Sequence 60, App1
21	76	2.7	450	9 US-09-876-889-279	Sequence 279, App
22	75	2.6	8788	10 US-09-764-877-2295	Sequence 2295, App
23	75	2.6	8788	10 US-09-860-670-247	Sequence 247, App
24	72.8	2.6	520	9 US-09-864-761-8048	Sequence 8048, App
25	60.2	2.1	65	12 US-09-908-975-24426	Sequence 24426, A
26	60	2.1	60	12 US-09-908-975-6184	Sequence 6184, App
27	54.2	1.9	500	12 US-09-814-353-12798	Sequence 12798, App
28	54.2	1.9	1559	12 US-09-814-353-19072	Sequence 19072, A
29	52	1.8	956	13 US-10-027-632-11508	Sequence 11508, A
30	51.6	1.8	1437	14 US-10-128-714-2188	Sequence 2188, App
31	51.6	1.8	1437	14 US-10-128-714-2188	Sequence 2188, App
32	50.6	1.8	2046	14 US-10-126-761-4538	Sequence 4538, App
33	50.6	1.8	9025608	14 US-10-156-761-1	Sequence 1, App1
34	50.4	1.8	152331	13 US-10-095-407-16	Sequence 16, App1
35	49.8	1.8	154746	12 US-09-827-688-8	Sequence 8, App1
36	49.4	1.7	10464	9 US-09-957-974-1	Sequence 1, App1
37	49	1.7	1041	9 US-09-916-790-3	Sequence 3, App1
38	49	1.7	1275	10 US-09-734-032-1	Sequence 1, App1
39	49	1.7	1275	13 US-10-016-985-1	Sequence 1, App1
40	49	1.7	1473	13 US-10-016-985-3	Sequence 3, App1
41	49	1.7	2598	9 US-09-816-094-1	Sequence 1, App1
42	49	1.7	2598	14 US-10-233-613-1	Sequence 1, App1
43	49	1.7	2893	9 US-09-916-790-1	Sequence 1, App1
44	49	1.7	7301	9 US-09-816-094-3	Sequence 3, App1
45	49	1.7	7301	10 US-09-734-032-3	Sequence 3, App1

ALIGNMENTS

RESULT 1
US-10-102-806-137
; Sequence 137, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 137
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,c,g, or c
US-10-102-806-137
Query Match 51.3%, Score 1457.4, DB 14, Length 2017,
Best Local Similarity 98.2%, Pred. No. 0;
Matches 1548; Conservative 8; Mismatches 10; Indels 10; Gaps 8;
QY 1272 ATGTCATAGTGTGACCCGATGTCAGAAAAGACGCTTGTATGACATTGATGTTG 1331
DB 4 ATGTCATAGTGTGACCCGATGTCAGAAAAGACGCTTGTATGACATTGATGTTG 63
QY 1332 AAGTGTATGACACCTTGAAGACCCAGATGATGATTTCTTCTGTCATGTCAGACCAAC 1391
DB 64 AAGTGTATGACACCTTGAAGACCCAGATGATGATTTCTTCTGTCATGTCAGACCAAC 123

QY 1392 AGGAGTTGCTACTCTAGACAAACAGCAATGACTGATGTGGTGGTAAACCCAGAGGAG 1451
DB 124 AGGAGATTGCTACTCTAGACAAACAGCAATGACTGATGTGGTGGTAA - CCARARGAG 182
QY 1452 AGCGCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGGAGGCTGTGGCGATCTTCT 1511
DB 183 AGCGCGAGCTGAGTTCTACTTCCAGCCCTGGG - KCAGGAGGCTGTGGCCRATCTTCT 241
QY 1512 ACTCCAAAGGTGACGACAGACAGCAAGAAATAGACAAGCCCTGGGAATCCGGAAATACAT 1571
DB 242 ACTCAAAGGTGACGACAGACAGCAAGAAATAGACAAGCCCTGGGAATCCGGNAATACAT 301
QY 1572 AGGGCTCTCTCCACAGCCCTGATTTCGACTGCACCAATTTCTTGAATTTGGGCCCTGTGCTGC 1631
DB 302 AGGGCTCTCTCCACAGCCCTGATTTCGACTGCACCAATTTCTTGAATTTGGGCCCTGTGCTGC 361
QY 1632 CTGCTCATAGPATCTGCTTGGTCTTGGTGGGGCGTTCCAGGGATGCTGTGGTTCA 1691
DB 362 CTGCTCATAGPATCTGCTTGGTCTTGGTGGGGCGTTCCAGGGATGCTGTGGTTCA 421
QY 1692 AGGCAAGACCAAGATGAAGGGTCTCACAAGACACCTGTTATCTCTTTTCCACCT 1751
DB 422 AGGCAAGACCAAGATGAAGGGTCTCACAAGACACCTGTTATCTCTTTTCCACCT 481
QY 1752 ATCTCTTCCACCCCTGCTTCCCTTTGCCCCACAAAGTTCCCATGTGCTGTACCCCTCC 1811
DB 482 ATCTCTTCCACCCCTGCTTCCCTTTGCCCCACAAAGTTCCCATGTGCTGTACCCCTCC 541
QY 1812 CTGCTCATAGPATCTGACTCTAGATAGTGTGTAGAGAGAACATGTAGTGTAAATGAGTG 1871
DB 542 CCTGCTCTACATAGGACCTCTAGATAGTGTGTAGAGAGAACATGTAGTGTAAATGAGTG 601
QY 1872 CTTGGAATGGATT - GSCCTCAGCCAGGCTGCTTCAAGGGACAGCAATCTGATCTCTA 1930
DB 602 CTTGGAATGGATTGGGCTCAGCCAGGCTGCTTCAAGGGACAGCAATCTGATCTCTG 661
QY 1931 CCCTTCAGAGACCCAGGAGTTGGG - - - TTTGCTCTCTTCTCCAAAGCTCAGGCTGTGGG 1987
DB 662 CCCTTCAGAGACCCAGGAGTTGGGAGCTTTGGCTCTCTTCTCCAAAGCTCAGGCTGTGGG 721
QY 1988 CACTCTATAAGTAGTTGATCTTGGCTCTCTGATATAAGCAATCCAAATTTCTCTCTTCC 2047
DB 722 CACTCTATAAGTAGTTGATCTTGGCTCTCTGATATAAGCAATCCAAATTTCTCTCTTCC 781
QY 2048 CTCACAGGTTGGAAACAACTCTCCCTTCACTCTGCTGCTGCTAGCACTACAGAAACC 2107
DB 782 CTCACAGGTTGGAAACAACTCTCCCTTCACTCTGCTGCTGCTAGCACTACAGAAACC 841
QY 2108 TGGTTCTT - GGCTCCACTGAGCCCCAGGTCAGTCCCCAG - CCTCTGGGTTGGCTGTGT 2165
DB 842 TGGTTCTTGGGCTCCACTGGCCCCAGGTCAGTCCCCAGCCCTCTGGGTTGRCCTGCTGT 901
QY 2166 CAGTCTCTCTCACTCTTGGTGGGTCACATCAGTATTTGGAGTTTGTCTTTATT 2225
DB 902 CAGTCTCTCTCACTCTTGGTGGGTCACATCAGTATTTGGAGTTTGTCTTTATT 961
QY 2226 GCTCCCTCCAGACACTCCCTGCGCTGCCCTTTGTGATTCCTCAGATCTGCCCTAATC 2285
DB 962 GCTCCCTCCAGACACTCCCTGCGCTGCCCTTTGTGATTCCTCAGATCTGCCCTAATC 1021
QY 2286 CGGGCATTTGGGTTGGGGAATCTTTGCCCTTTCCCTTTTCAGAGCCCCAGGATCTCATCTG 2345
DB 1022 CGGGCATTTGGGTTGGGGAATCTTTGCCCTTTCCCTTTTCAGAGCCCCAGGATCTCATCTG 1081
QY 2346 GGGAACTGTTCATTTGCACAGAGGCTGTTCCTCTCTGCAAGTTTGGAGATGTGACTCATTC 2405
DB 1082 GGGAACTGTTCATTTGCACAGAGGCTGTTCCTCTCTGCAAGTTTGGAGATGTGACTCATTC 1140
QY 2406 CATTCACCTACCTCCACCTGCTCTGCACTCCCTTAATGGAGAAA CGGGCCTAAACCAA 2465
DB 1141 CATTCACCTACCTCCACCTGCTCTGCACTCCCTTAATGGAGAAA CGGGCCTAAACCAA 1200

QY 2466 CGGTAATAAAGCCCTGGGCCATCCCTGTCTTCTCTGTCTCTTGTCTGCCAGTTGACACC 2525
DB 1201 CGGCT - AAAAAGCCCTGGGCCATCCCTGTCTTCTCTGTCTCTTGTCTGCCAGTTGACACC 1259
QY 2526 TACTGGTGACTCTTAGGGCACTAGAGGAGTGAAGCGCTAGGGCTGGAGAAATAGCGCTGA 2585
DB 1260 TACTGGTGACTCTTAGGGCACTAGAGGAGTGAAGCGCTAGGGCTGGAGAAATAGCGCTGA 1319
QY 2586 GTTGGGTTGTGACTCTTCTCTCTCTCCCTCAGAGGATGTGACTCTCCCGAGCCCTTGC 2645
DB 1320 GTTGGGTTGTGACTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1379
QY 2646 CCTCAAAGCTTCAGACCCCTCAGTAGCAGAGACCTTGTGATCTTGGCCCTTGGATC 2705
DB 1380 CCTCAAAGCTTCAGACCCCTCAGTAGCAGAGACCTTGTGATCTTGGCCCTTGGATC 1439
QY 2706 TGAGATGGTTTTTGCATCTTTCCAGGAGAGCCTCACAATCTTCTTCCAGGTTGTATCACC 2765
DB 1440 TGAGATGGTTTTTGCATCTTTCCAGGAGAGCCTCACAATCTTCTTCCAGGTTGTATCACC 1499
QY 2766 CCCAGTTAGCATATCCAGGCTCGCAGACTCAACACAGAAAGGTTGGAGACAGCTGGG 2825
DB 1500 CCCAGTTAGCATATCCAGGCTCGCAGACTCAACACAGAAAGGTTGGAGACAGCTGGG 1559
QY 2826 CACAAAGGGGATTC 2841
DB 1560 CACAAAGGGGAATTC 1575

RESULT 2
US-09-920-300A-11/c
; Sequence 11, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Mesgher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920.300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 365_433, 507, 549, 591
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-11

Query Match 17.2%; Score 488.8; DB 10; Length 606;
Best Local Similarity 93.6%; Pred. No. 8.7e-144;
Matches 553; Conservative 0; Mismatches 30; Indels 8; Gaps 4;
QY 2237 GACACTCCCTGTGGCTGCC - - - TTTGTGATTCCTCTCAGATCTGCC - - - TAAATCCCGG 2290
DB 590 GACACTCCCTGTGGCTGCCCTTTGTGAAATTCCTCTCAGATNTGCCCTTAATCCCGGG 531
QY 2291 CATTTGGGTTGGGAATCTTGGCTTCCCTTCAGAGCCCGAGGATCTCATCTGGGAA 2350
DB 530 CATTTGGGTTGGGAATCTTGGCTTCCCTTCAGAGCCCGAGGATCTCATCTGGGAA 471
QY 2351 CTGTCAATGCCAGAGGCTGTTCCTTCTCTGCAAGTT - TGGAGATGTGACTCATTCATTT 2409
DB 470 CTGTCAATGCCAGAGGCTGTTCCTTCTCTGCTGTTTGGAGATGTGACTCATTCATTC 411
QY 2410 CACTCACTCCACCTGCTCTGCAATCCCTTAATGGAGAAA CGGGCCTAAACCAAACGGG 2469

Db 410 ACTCACTCCACCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 351
 Qy 2470 TAAAAAGCCCTGGGCAATCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 2529
 Db 350 T-AAAAAGCCCTGGGCAATCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 292
 Qy 2530 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 2589
 Db 291 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 232
 Qy 2590 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 2649
 Db 231 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 172
 Qy 2650 AAAGCTTACAGCCCTCAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 2709
 Db 171 AAAGCTTACAGCCCTCAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 112
 Qy 2710 ATGATTTTGCATCTTTCCAGAGAGCTCAGATTTCTTCCAGAGTTGATCAGCCCG 2769
 Db 111 ATGATTTTGCATCTTTCCAGAGAGCTCAGATTTCTTCCAGAGTTGATCAGCCCG 52
 Qy 2770 AGTTAGCATATCCAGAGCTCGAGACTCAACAGCAAGGGTGGAGACAG 2820
 Db 51 AGTTAGCATATCCAGAGCTCGAGACTCAACAGCAAGGGTGGAGACAG 1

RESULT 3
 US-10-099-926-11/c
 ; Sequence 11, Application US/10099926
 ; Publication No. US20030166064A1
 ; GENERAL INFORMATION:
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Meagher, Madeleine Joy
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Jiang, Yudi
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.547C2
 ; CURRENT FILING DATE: 2002-03-17
 ; NUMBER OF SEQ ID NOS: 1982
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 606
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 365, 433, 507, 549, 591
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-099-926-11

Query Match 17.2%; Score 488.8; DB 12; Length 606;
 Best Local Similarity 93.6%; Pred. No. 8,7e-144;
 Matches 553; Conservative 0; Mismatches 30; Indels 8; Gaps 4;
 Qy 2237 GACACTCCCTGCTGCTGCC--TTTGTGATTCCTCAGATCTGCC--TAATCCGGG 2290
 Db 590 GACACTCCCTGCTGCTGCC--TTTGTGATTCCTCAGATCTGCC--TAATCCGGG 531
 Qy 2291 CATTTGGGGGGGGAATCTTGCCTTTCCCTTCAAGAGGATCTCAATCTGGGAA 2350
 Db 530 CATTTGGGGGGGGAATCTTGCCTTTCCCTTCAAGAGGATCTCAATCTGGGAA 471
 Qy 2351 CTGTATTTGCCAGAGAGGCTGTTCTTCTCTGAGATT-TGAGATGTGACTCATTTCAAT 2409
 Db 470 CTGTATTTGCCAGAGAGGCTGTTCTTCTCTGAGATT-TGAGATGTGACTCATTTCAAT 411
 Qy 2410 CACTCACTCAACCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 2469
 Db 410 ACTCACTCAACCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 351

Qy 2470 TAAAAAGCCCTGGGCAATCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 2529
 Db 350 T-AAAAAGCCCTGGGCAATCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 292
 Qy 2530 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 2589
 Db 291 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 232
 Qy 2590 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 2649
 Db 231 GGTGACTTCTAGGCACTGAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 172
 Qy 2650 AAAGCTTACAGCCCTCAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 2709
 Db 171 AAAGCTTACAGCCCTCAGAGAGTAAAGCGGCTGAGAAATAGCGCTGAGTTG 112
 Qy 2710 ATGATTTTGCATCTTTCCAGAGAGCTCAGATTTCTTCCAGAGTTGATCAGCCCG 2769
 Db 111 ATGATTTTGCATCTTTCCAGAGAGCTCAGATTTCTTCCAGAGTTGATCAGCCCG 52
 Qy 2770 AGTTAGCATATCCAGAGCTCGAGACTCAACAGCAAGGGTGGAGACAG 2820
 Db 51 AGTTAGCATATCCAGAGCTCGAGACTCAACAGCAAGGGTGGAGACAG 1

RESULT 4
 US-10-033-528-11/c
 ; Sequence 11, Application US/10033528
 ; Publication No. US20020131971A1
 ; GENERAL INFORMATION:
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Meagher, Madeleine Joy
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Secrist, Heather
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.547C1
 ; CURRENT FILING DATE: 2001-12-26
 ; NUMBER OF SEQ ID NOS: 1896
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 606
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 365, 433, 507, 549, 591
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-033-528-11

Query Match 17.2%; Score 488.8; DB 13; Length 606;
 Best Local Similarity 93.6%; Pred. No. 8,7e-144;
 Matches 553; Conservative 0; Mismatches 30; Indels 8; Gaps 4;
 Qy 2237 GACACTCCCTGCTGCTGCC--TTTGTGATTCCTCAGATCTGCC--TAATCCGGG 2290
 Db 590 GACACTCCCTGCTGCTGCC--TTTGTGATTCCTCAGATCTGCC--TAATCCGGG 531
 Qy 2291 CATTTGGGGGGGGAATCTTGCCTTTCCCTTCAAGAGGATCTCAATCTGGGAA 2350
 Db 530 CATTTGGGGGGGGAATCTTGCCTTTCCCTTCAAGAGGATCTCAATCTGGGAA 471
 Qy 2351 CTGTATTTGCCAGAGAGGCTGTTCTTCTCTGAGATT-TGAGATGTGACTCATTTCAAT 2409
 Db 470 CTGTATTTGCCAGAGAGGCTGTTCTTCTCTGAGATT-TGAGATGTGACTCATTTCAAT 411
 Qy 2410 CACTCACTCAACCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 2469
 Db 410 ACTCACTCAACCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 351
 Qy 2470 TAAAAAGCCCTGGGCAATCCCTGCTCTGATCCCTTAATGAGAAAAGGGGCTAAACCAAGCGG 2529

Db 350 T-AAAAAGCCCTGGGCAATCCCTGTCTTCTGCTCCCTTGTCTGCCAGTTGACACCTACT 292
QY 2530 GGTGACTTCTAGGCACTGAGAGTGAAGCGCTAGGCTGGAGAAATAGCGCTGAGTTG 2589
Db 291 GGTGACTTCTAGGCACTGAGAGTGAAGCGCTAGGCTGAAGAAATAGCGCTGAGTTG 232
QY 2590 GGTGTTGTGACTCTTCCCTCTCCCTGCTCAGAGGATTTGACTCCCCAGCCCTGCCCTC 2649
Db 231 GGTGTTGTGACTCTTCCCTCTCCCTGCTCAGAGGATTTGACTCCCCAGCCCTGCCCTC 172
QY 2650 AAAGCTTTCAGACCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGCCCTTGGATCTGAG 2709
Db 171 AAAGCTTTCAGACCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGCCCTTGGATCTGAG 112
QY 2710 ATGGTTTGTGATCTTTCAGAGAGCCTCAGATCTTCTTCCAGGTTGTATCACCCTCG 2769
Db 111 ATGGTTTGTGATCTTTCAGAGAGCCTCAGATCTTCTTCCAGGTTGTATCACCCTCG 52
QY 2770 AGTTAGCATATCCAGGCTCGAGACTCAACAGCAAGGGTGGAGACAG 2820
Db 51 AGTTAGCATATCCAGGCTCGAGACTCAACAGCAAGGGTGGAGACAG 1

RESULT 5
US-09-879-536-796/c
; Sequence 796, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 796
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-796

Query Match 13.7%; Score 388.8; DB 10; Length 452;
Best Local Similarity 97.9%; Pred. No. 3.6e-112;
Matches 415; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
QY 566 CTGGTACAGAAATCCAGCCCTATATGGATCTCTTGGCTTTTGAAGAAACTGACACAG 625
Db 423 CAGGTACAGAAATCCAGCCCTATATGGATCTCTTGGCTTTTGAAGG-AACTGGACACAG 365
QY 626 ACTATCATCAGAAACGGCTAGATATCAAGAGGCTTGAACGCTCCCTTAAGCAAAA 685
Db 364 ACTATCATGAGG-AACGGCTAGATATCAAGAGGCTTGAACGCTCCCTTAAGCAAAA 306
QY 686 CGAAGCTCGCAATTTTTCATTTTCAACTCCGGCTAAGTCAGATGCCAGGAT 745

Db 305 CGGAAGCTGCGAATTTTTCATTTTCAACTTCGCGCTAAGTCAGATGCGAGGAT 246
QY 746 GGGGAAGGGAGCGGTGGCTTCTGGAGCTTCGGGTAGAGGACGGCTCCTCGAGATTCA 805
Db 245 GGGGAAGGGAGCGGTGGCTTCTGGAGCTTCGGGTAGAGGACGGCTCCTCGAGATTCA 186
QY 806 GCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGGAAGTTCTCTTCTTTAAAGTCC 865
Db 185 GCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGGAAGTTCTCTTCTTTAAATCC 126
QY 866 TTGTTGATTAATGGAACAAAGACCTGTATGGGCCAGACCAATCTGTTAGATGGCAC 925
Db 125 TTGTTGATTAATGGAACAAAGACCTGTATGGGCCAGACCAATCTGTTAGATGGCAC 66
QY 926 AGGACCCCACTACCCAGGAGACCGATGGCTTCAGGTGAAGCGCCAGGAGATGTGAAT 985
Db 65 AGGACCCCACTACCCAGGAGACCGATGGCTTCAGGTGAAGCGCCAGGAGATGTGAAT 6
QY 986 GTAC 989
Db 5 GTAC 2

RESULT 6
US-09-879-536-645/c
; Sequence 645, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 645
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(690)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-645

Query Match 10.9%; Score 310.2; DB 10; Length 690;
Best Local Similarity 99.0%; Pred. No. 3.8e-87;
Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1409 GACAACAAGCAATGACTGATGCTGGGTAAACCCAGAGGAGGAGCGCCGAGCTGAGTTC 1468
Db 315 GACCTCAAGCAATGACTGATGCTGGGTAAACCCAGAGGAGGAGCGCCGAGCTGAGTTC 256
QY 1469 TACTTCCAGCCCTGGGCTCAGGAGGCTGTGCGGATCTTCTACTCCAAAGTCGACAG 1528
Db 255 TACTTCCAGCCCTGGGCTCAGGAGGCTGTGCGGATCTTCTACTCCAAAGTCGACAG 196
QY 1529 AGACACAAGCAATTTAGAGCAAGCCCTGGGAATCCGGAATACATAGGGCCCTCTCCACAGC 1588

```
Db      195 AGACGACAAAGATTAGAGCAGCCCTGGGAAATCCGGAATCATAGGAGCCCTCCACAGC 136
QY      1589 CCGATTTCAGCTGCACCAATTTCTGATTGGGCGCGTGTGCTCCCTCATAGTATCG 1648
Db      135 CCGATTTCAGCTGCACCAATTTCTGATTGGGCGCGTGTGCTCCCTCATAGTATCG 76
QY      1649 CCTGTGCTTGTGTTGGGCGTTCCAGGGAGTGTGTTGTTCAAGACAAGCCAGATG 1708
Db      75 CCTGTGCTTGTGTTGGGCGTTCCAGGGAGTGTGTTGTTCAAGACAAGCCAGATG 16
QY      1709 AAGAGGCTCTCAAA 1723
Db      15 AAGAGGCTCTCAAA 1

RESULT 7
US-09-918-995-16164
; Sequence 16164, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16164
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(485)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-16164

Query Match
Best Local Similarity 73.4%; Pred. No. 1,6e-72;
Matches 337; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY      819 ATGATGCCACTAAACAAAGAGAGTTCTTCTCTTTTAAGTCTTGGTGAATGAAC 878
Db      26 AGGAATCTCAGACAGACAGAGCGGAGTTCTTCTTTTCAAGATTGTGTCATGAGC 85
QY      879 TGGACAAACCTGTATTTGGGCCAGACCAACCATTTGTTAGATGGCACAGAGCCGACCTA 938
Db      86 TGGACAAACCTTTTATTTGGCCCTGACCAACCATCTGTTAGTGGCACTGGACACCCACGA 145
QY      939 CCAGAGAGACCGATGCTTCCAGAGTGAAGCGCCAGAGATGTAATGAATGACGATGTAAC 998
Db      146 CCCAGAGAGACGAGCGCTTCCAGGTGAAGCGCTTGGGACCTGAGATGTGCGCTGACCC 205
QY      999 TCTTACTGATGCTGATTAACAGCTCCCAAGTTTAATTAAGACCCCGCTTACGTGAC 1058
Db      206 TGTCTCTCATGTGCTGACTACAGCTCCCAAGTTTAACTGATATCCCGCTTAAACCGGC 265
QY      1059 TCTTGGGACATCCATACCCAGACCTCGCTCCAGTATCATCCAGAGACTGTGGCATATATTA 1118
Db      266 TGTGTGGGCTGCACACAGAGCCGCTCAGCCATTGTCCAGGCGCTGTGGCAGTATGTA 325
QY      1119 AGACATTAAGCTCCAGACCTCTCAGAGCGGAGTTTGTTCATCTGTGACAGTACCTGC 1178
Db      326 AGACCAACAGCTGCGACGATCCCATGACAGGAATATCATGAGGAGCAAGTATTTCC 385
QY      1179 AGCAGATCTTTGAGACTCAACGATGTAAGTTTTCAGAGATCCCTCAGCGGCTTCATGCT 1238
Db      386 AGCAGATGTTTGAATGTCTCCCGGCTGAAGTTTCTGAGATTACCCAGCGCTCACAGCCC 445
```

```
QY      1239 TGTATTGCCACAGAACCTATCATCATTAATCATGTCA 1277
Db      446 TGTATTGCCCGCTGACCCCAATTGTTCATCAACCATGTCA 484

RESULT 8
US-09-918-995-26465
; Sequence 26465, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ-ID NO 26465
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)....(474)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26465

Query Match
Best Local Similarity 72.5%; Pred. No. 2.7e-64;
Matches 308; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY      899 CCAGACCAACCATCTGTGTAATGAGCAGAGACCGCCACTACCCAGAGACCGATGCTTC 958
Db      50 CCTGACCAACCACTGTTGAGTGCGATCGACACCCACGACGAGACGAGCGGCTTC 109
QY      959 CAGGTGAACCGCCAGAGATGTAATGATGAGTACTGTCTACTGATGCTGATTAAC 1018
Db      110 CAGGTGAACCGGCTGGGGACCTGAGTGTGCGCTCAGCGTCTCTCATGTGACTAC 169
QY      1019 CAGCCTCCCAATTAATTAATAGACCCCGCTAGCTGACCTCGGAGATCCATACCCAG 1078
Db      170 CAGCCTCCCAATTAATAGACTGATCCCGCTATCCCGCTGTGGGCTGCACACACAG 229
QY      1079 ACTGTCCAGTATCATCCAGACCTGTGGCAATATTAAGACAATAGCTCCAGAGAC 1138
Db      230 ACCGGTACGCCATTGTCCAGGCCCTGTGGCAGTATGTAAGACCAACAGCTGCAGAGAC 289
QY      1139 CCTCAGACGGGAGTTGTCTATCTGTGACAAATGCTGACAGAGATCTTTGAGACTCA 1198
Db      290 TCCCATGACAGAGAAATACATCAATGAGGACAAGTATTTTCCAGCAATTTTGTATGCC 349
QY      1199 CGTATGAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTATGCAACAGAACT 1258
Db      350 CGGCTGAGTTTTCAGAGTTTACCCAGCGCTTCAAGACCTGTCTATTTGCCCTTACCCA 409
QY      1259 ATCATCATTAATCATGATCATAGTGTGACCCGAAATGATCAAGAAAGACAGCTTGTAT 1318
Db      410 ATTGTCAATCAATCATGATCATAGCGGTGACCTTTCAGACAGAAAGACGCGGTGTAT 469
QY      1319 GACAT 1323
Db      470 GACAT 474

RESULT 9
US-09-918-995-28419
; Sequence 28419, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
```

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28419
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(447)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28419

Query Match 7.8%; Score 222.6; DB 11; Length 447;
Best Local Similarity 74.1%; Pred. No. 1.7e-59;
Matches 297; Conservative 0; Mismatches 99; Indels 5; Gaps 1;
QY 583 GGCCTATATGGATCTCTTGGCTTTTGAAGGAAACTGGACACAGACTATCATGAGGAAACG 642
Db 52 GGAATGCATGGACCTCTTGGCATTTTGAAGGAAACTGGATCAAAACCATCATGCGGAAGCG 111
QY 643 GCTAGATATCAAGAGCGCTTCAAGCTCCCAATTAAAGCAAAACGGAAGCTCGCAATTTT 702
Db 112 GGTGACATCCAGGAGGCTCTGAAGAGGCCCAATGAAGCAAAAGCGGAAGCTGCGACTCTA 171
QY 703 CATTTCTTAACATTTTCAATCCGGCTAAGTCAGATCCGAGGATGGGAAGGACGCGTGGC 762
Db 172 TATCTCCACACATTTTAAACCTTGCAGGCTGATGCTGAGATTTCGACGCGCAGATTGC 231
QY 763 TTCTCTGGAGCTTCGGGTAGAGGACGCTCTCGGAGGATTCAGCTTGTCCAAATATGA 822
Db 232 CTCTGGGAGCTACGGGTGGAGGGAAGCTCTCGATGAT-----GTACGTCCTCGGCCCA 286
QY 823 TGCCACTAAACAAAGAGGAAGTTCTCTCTCTTCTTAAAGTCTTGGTGAATGAATCTGA 882
Db 287 GCCCAGCAACAGAGCGGAAGTTCTCTCTTCTTCAAGAGTTTGGTCAATGAGCTGA 346
QY 883 CAAAGACCTGTATGGCGCAGACAAACATCTGTAGATGGCAGGACCGCCACTACCCA 942
Db 347 CAAAGATCTTATGGCCCTGACACACCACTCTGTTAGTGGCATCGACACCCACGCCCA 406
QY 943 GGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGATGTGA 983
Db 407 GGAGAGGACGCGCTTCCAGGTCAAACGCGCTGGGACCTGA 447

RESULT 10
US-09-918-995-2832
; Sequence 2832, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Bracco, Laurent
; APPLICANT: Schweighoffer, Fabien
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2832
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(495)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2832
Query Match 6.1%; Score 174.4; DB 11; Length 495;
Best Local Similarity 65.4%; Pred. No. 3.4e-44;
Matches 291; Conservative 0; Mismatches 142; Indels 12; Gaps 2;
QY 666 AACGTCCCATTAAGCAAAACGGAAGCTGCGAAATTTTCAATTTCTAAACACTTTCAATCCGG 725
Db 50 AAAAGCCTCTGACACAAAAGCGAAGCTTCGGATCTACATTTCCAAATCGTTCAGTCCCA 109
QY 726 CTAAGTCAGATGCCGAGGATGGGAAGGACGGTGGCTTCTGGGAGCTTCGGGTAGAAG 785
Db 110 GCAAGCGGAAGCGCATAGTCAGGAACTGCAGGACCCCTGGGGGAAACCCAGCAGGGG 169
QY 786 ----GACGGCTCTGGAGGATTAGCCCTTGTCCAAA-----TATGATGCCACTAAAC 833
Db 170 ACAAGGTGGCTTCTCGGAACTCCGAGTGGGAAGAAAACCTGCTGGATGATCTTAGCAAC 229
QY 834 AAAAGAGGAAGTCTCTTCTTCTTCTTAAAGTCTTTGGTGAATGAACCTGGACAAAGACCTGT 893
Db 230 AGAAGAGGAAGTCTTCTTCTTCTTAAAGACCTCGTCAATTCAGCTGGCAAGGAGCTGT 289
QY 894 ATGGGCCAGACAACCATCTGGTAGAATGGCAGGACCGCCACTACCCAGAGACCGATG 953
Db 290 ACGGCGCTGACAATCACTCCTGGTGGAGTGGCACCAGATGCCACCACCCAGGAGACAGATG 349
QY 954 GCTTCCAGGTGAAGCGGCGCAGGAGATGATGATGATGATGATGATGATGATGATGATGATG 1013
Db 350 GCTTCCAAGTAAAAACGCGCTGGAGACCTCAAAGTCAAGTGCACCTCTCTGCTCATGCTGG 409
QY 1014 ATTACGAGCTCCCGAGTTTAAATTTAGACCCCGCCCTAGCTCGACTCTCTGGGCATCCATA 1073
Db 410 ATCATGACCTTCCCGAGTACAAATTTGACCCCGCCGATTTGGCAAGGCTTGTCTGGAGTGCACA 469
QY 1074 CCCAGACTCGTCCAGTGCATCATCCA 1098
Db 470 CGCAGACAAAGGCGCNCATCATGCA 494

RESULT 11
US-10-070-676-23/c
; Sequence 23, Application US/10070676
; Publication No. US20030059788A1
; GENERAL INFORMATION:
; APPLICANT: Toque, Bruno
; APPLICANT: Bracco, Laurent
; APPLICANT: Schweighoffer, Fabien
; TITLE OF INVENTION: Genetic Markers of Toxicity, Preparation
; TITLE OF INVENTION: and Uses
; FILE REFERENCE: 50146/003002
; CURRENT APPLICATION NUMBER: US/10/070,676
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: PCT/FR00/02503
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: FR 99/11405
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/456,370
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3, 33, 34
; OTHER INFORMATION: n = A,T,C or G
US-10-070-676-23

Query Match 4.3%; Score 123.4; DB 14; Length 310;

Best Local Similarity 66.0%; Pred. No. 3.7e-28;
Matches 192; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

QY 1097 CAAGCACTGCAATATATTAGACACATTAAGCTCCAGAACCTCCAGAGGGGGGAGTT 1156-
DB 306 CAGGCGCTTGCTTTTACATCAAGCAACACAGCTCAGATGGGACAGCGGGATAC 247
QY 1157 GTCATCTGACAGTACCTGACAGATCTTTGAGACTCAAGTATGAAGTTTTCAGAG 1216
DB 246 ATCACTGACACCGCTTACTTCCGCCAGATC-TTCACTTGGGCGCATCCGTTTCTCCGAG 188
QY 1217 ATCCCTCAGCGGCTCCATGCTTGTATGACACAGAACTATCATTAATCATGTC 1276
DB 187 ATCCCATGAACTGCGAGGTTGTGACAGATCCAGACCCCATTTGTATCAACCATGTC 128
QY 1277 ATCAGTGTGACCCCAATATATCAAGAAAAGACGCTTTTATGACATGATGTGAAGT 1336
DB 127 ATTAGGTGACCTTAACGACGAGAAAGACGCTGTTTACGACATGATGTGAGAGT 68
QY 1337 GATGACACCTTGAAAGACGAGATGATCTTTTCTGCTGCTCCAGCCAGC 1387
DB 67 GACGACCCACTGAAGGCCCAATAGACATTTTNGGCTCTACACCATC 17

RESULT 12
US-09-960-352-12390
Sequence 12390, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Waiten, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 12390
LENGTH: 285
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 53-LIB3058-001-Q1-K1-F2
US-09-960-352-12390

Query Match 4.0%; Score 113.6; DB 10; Length 285;
Best Local Similarity 65.8%; Pred. No. 4.5e-25;
Matches 192; Conservative 0; Mismatches 79; Indels 21; Gaps 1;

QY 656 GAGGCTTGAAAGCTCCCATTAAGCAAAAACGGAAGCTGCAATTTCTTCTAACACT 715
DB 15 GATGCTCTCAAGCTGTTTTCAGAAAGCGCAAGCTGCTGCTATATCTCCATTA 74
QY 716 TTCAATCCGGCTAAGTCAGATGCGGAGATGGGGAAGGGGCTGCTCCGGGAGCTT 775
DB 75 TTAACTCTGCTTAAGCTCCGATGCTTGAAGGAGCATCTCTCTCGGGAGCTG 134
QY 776 CGGCTGAAGAGAGCGCTCTGAGATTCAGCTTGTCCAAATATGATGCACTAAACA 835
DB 135 CTGGTGAAGAGGAGGAGCTCTGG-----ATACCCAGCTGAGCAT 173
QY 836 AAGAGGAAGTTCTCTTCTTTTAAAGTCTTGATGATGAACTGACAAAGCTGTAT 895
DB 174 ATGCGGAAGTTCTGCTCTTCAATCAAGAGCTCGTATGATCTGACATAGACCTTTAT 233
QY 896 GGGCCGAGCAACATCTGATAGATGCAAGGACCGGCACTAATCCAGAGAGA 947
DB 234 GGCCCTGACACCACTGATGAGTGGACCTGACACCACTACAGAGA 285

RESULT 13
US-09-783-590-3701

/ Sequence 3701, Application US/09783590
/ Patent No. US20020110850A1
/ GENERAL INFORMATION:
/ APPLICANT: Dillion, Patrick J.
/ APPLICANT: Haseltine, William A.
/ APPLICANT: Li, Haodong
/ APPLICANT: Rosen, Craig A.
/ APPLICANT: Ruben, Steven M.
/ TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
/ FILE REFERENCE: PO-16.2C1
/ CURRENT APPLICATION NUMBER: US/09/783,590
/ CURRENT FILING DATE: 2000-02-15
/ PRIOR APPLICATION NUMBER: 08/420,856
/ PRIOR FILING DATE: 1995-04-12
/ PRIOR APPLICATION NUMBER: 08/346,731
/ PRIOR FILING DATE: 1994-11-21
/ NUMBER OF SEQ ID NOS: 12485
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 3701
/ LENGTH: 356
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (5)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (8)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (12)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (23)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (44)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (92)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (120)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (138)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (237)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (248)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (252)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (268)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (314)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (325)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (327)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (331)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (347)

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (353)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3701

Query Match      3.7%; Score 104; DB 10; Length 356;
Best Local Similarity 88.9%; Pred. No. 5.7e-22;
Matches 200; Conservative 0; Mismatches 15; Indels 10; Gaps 8;

QY 1416 ACACAAATGACTGATGTGTGGTACCCAGGAGGAGCGCGGAGCT-GAGTTCTACTTC 1474
Db 22 ANACAATGACTGATGTGTGGGNAACCCAAAGGAGGAGCGCGGAGCTGGAGTTCTACTTC 81

QY 1475 CAGCCCTGGGCTCAGG-AGGCTGTGTGGGATCTTCT-ATCCAAAGTGAGCAGAGAGA- 1531
Db 82 CAGCCCTGGGNTCAGAAAGGCTGTGTGGGATCTTCTNACTCCAAAGTGAGCAGAGNGAA 141

QY 1532 -CGACAAGAATTAGACGAAGCCCTGGGAA--TCCGGAATACATAGGGCCTCT-CCACAG 1587
Db 142 CGACAAGGAATTAGACGAAGCCCTGGGAAATCCGGTAATACATAGGGCCTCTGCCACAG 201

QY 1588 CCTGATTTCGACT-GCACAATTCTT-GATTTGGCCCTGTGCTG 1630
Db 202 CCTGATTTCGACTGGCACCATTCTTGGATTGGGNCCTGTGATG 246

RESULT 14
US-09-918-995-34033
; Sequence 34033, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34033
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34033

Query Match      3.6%; Score 102.6; DB 11; Length 420;
Best Local Similarity 71.4%; Pred. No. 1.8e-21;
Matches 135; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1009 GCTGATTACAGGCTCCCGAGTTTAAATAGACCCCGCTAGCTCAGCTCTGGGCAT 1068
Db 3 GCTGACTACCAAGCTCCCGAGTTTCAAACTGGATCCCGCTAGCCGCTGCTGGGCT 62

QY 1069 CCATACCCAGACTCTCCAGTGATCATCCAAGCACTGTGGCAATATATTAGACACATAA 1128
Db 63 GCACACACAGAGCCCTCAGCCATTGTCCAGCCCTGTGGAGTATGTAGAGACCAACAG 122

QY 1129 GCTCCAGGACCCCTACGAGCGGGAGTTTGTTCATCTGTGCAAGTACCTGCAGCAGATCTT 1188
Db 123 GCTGCAGGACTCCCATGACAGGAATACATCAATGGGACAAGTATTTCCAGCAGATTCC 182

QY 1189 TGAGACTCA 1197
Db 183 TGAGTAGCA 191

RESULT 15
US-09-918-995-28488
; Sequence 28488, Application US/09918995
; Publication No. US20030073623A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28488
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(462)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28488

Query Match      3.4%; Score 97.4; DB 11; Length 462;
Best Local Similarity 73.1%; Pred. No. 8.3e-20;
Matches 125; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1409 GACAAACAAGACAATGACTGATGTGGTAAACCCAGAGGAGGAGCGCCGAGCTGAGTTT 1468
Db 280 GACCTCAAGGTGATGACAGATGTAGCCGCAACCCCTGAAGAGGAGCGCCGCTGAGTTT 339

QY 1469 TACTTCCAGCCCTGGGCTCAGAGGCTGTGTGCGATACTTCTACTCAAGGTGCAGAG 1528
Db 340 TACCACCAGCCCTGGTCCCGAGGCGCTCAGTCGCTACTTCTACTCAAGATCCAGCAG 399

QY 1529 AGACGACAAGAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGCGCTC 1579
Db 400 CGCAGGAGGAGCTGGAGCAGTCGCTGGTTGTGGCAACACCTTAGGAGCCC 450

Search completed: October 11, 2003, 17:35:10
Job time : 532 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 12:21:42 : Search time 3873 Seconds
(without alignments)
17828.303 Million cell updates/sec

Title: U66617

Perfect score: 2841
Sequence: 1 GAATTCGGCTATCCCATG.....TGGCACAAGGGGATTTC 2841

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estbda:*
2: em_estbnum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_ping:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2439.4	85.9	3206	11	BC000063 Homo sapi
2	892	31.4	929	13	BQ933801 AGENCOURT
3	847.8	29.8	933	13	BX386559 BX386559
4	839.4	29.5	1201	13	BX375270 BX375270

5	833.6	29.3	907	13	BQ929257	BQ929257 AGENCOURT
6	825.2	29.0	970	9	AL521992	AL521992 AGENCOURT
7	810.6	28.5	979	13	BQ934447	BQ934447 AGENCOURT
8	809.2	28.5	953	13	BQ845642	BQ845642 AGENCOURT
9	795.6	28.0	2476	11	AK075611	AK075611 Mus muscu
10	779.8	27.4	972	13	BQ881274	BQ881274 AGENCOURT
11	772.4	27.2	872	13	BQ931457	BQ931457 AGENCOURT
12	771.8	27.2	1003	12	BM457092	BM457092 AGENCOURT
13	765	26.9	1201	13	BX363764	BX363764 BX363764
14	763.6	26.9	858	13	BQ557139	BQ557139 AGENCOURT
15	760.2	26.8	878	13	BQ949766	BQ949766 AGENCOURT
16	755.2	26.6	820	9	AU132848	AU132848 AGENCOURT
17	755.2	26.6	1123	12	BQ061434	BQ061434 AGENCOURT
18	748.6	26.3	953	13	BQ716211	BQ716211 AGENCOURT
19	747.6	26.3	825	13	BQ521876	BQ521876 AGENCOURT
20	746.6	26.3	783	9	AU132128	AU132128 AGENCOURT
21	742	26.1	1094	13	BX340735	BX340735 BX340735
22	734	25.8	913	13	BQ591331	BQ591331 AGENCOURT
23	725.6	25.5	882	13	BQ232118	BQ232118 AGENCOURT
24	720.6	25.4	943	13	BQ927245	BQ927245 AGENCOURT
25	703.6	24.8	758	12	BI917156	BI917156 AGENCOURT
26	701.4	24.7	919	13	BQ652596	BQ652596 AGENCOURT
27	697.8	24.6	905	13	BX390897	BX390897 BX390897
28	696.8	24.5	781	10	BQ676290	BQ676290 AGENCOURT
29	695.4	24.5	914	13	BQ957208	BQ957208 AGENCOURT
30	694.8	24.5	917	13	BQ509684	BQ509684 AGENCOURT
31	690.8	24.3	785	13	BQ053012	BQ053012 UI-M-FCO-
32	688.4	24.2	1061	12	BM543298	BM543298 AGENCOURT
33	677.2	23.8	704	9	AU131948	AU131948 AGENCOURT
34	674	23.7	828	14	CA315273	CA315273 UI-M-FW0-
35	674	23.7	912	12	BI685523	BI685523 AGENCOURT
36	665.8	23.4	754	14	CA315141	CA315141 UI-M-FW0-
37	663.6	23.4	750	13	BQ571495	BQ571495 UI-M-FCO-
38	663.2	23.3	788	14	CA318066	CA318066 UI-M-FW0-
39	661.8	23.3	908	13	BQ422607	BQ422607 AGENCOURT
40	654.2	23.0	744	12	BI184681	BI184681 UNL-P-FN-
41	653.2	23.0	767	13	BQ613510	BQ613510 UI-M-EW0-
42	651.2	22.9	745	14	CA451068	CA451068 UI-M-FW0-
43	648.4	22.8	847	9	AU129746	AU129746 AGENCOURT
44	643.8	22.7	944	13	BQ919072	BQ919072 AGENCOURT
45	639.2	22.5	755	14	CA315454	CA315454 UI-M-FW0-

ALIGNMENTS

RESULT 1
LOCUS BC000063
DEFINITION Homo sapiens, Similar to SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1, clone IMAGE:3508933, mRNA.
ACCESSION BC000063
VERSION BC000063.1 GI:12652634
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3206)
TITLES Direct Submission
JOURNAL Submitted (03-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology

http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 7 Row: b Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4507082
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1..3206
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3508933"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH MGC 19"
/lab_host="DH10B-R"
/note="Vector: pOTB7" 772 t

BASE COUNT 738 a 913 c 783 g 772 t

ORIGIN

Query Match 85.9%; Score 2439.4; DB 11; Length 3206;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 2669; Conservative 0; Mismatches 11; Indels 133; Gaps 9;

QY 160 GGGCGGGTTTCAGTCTGTGGCTCCAAAGGGGGGGCGCGGAGCCCTCAGAGGGGGGGCGG 219
Db 9 GGGCGGGTTTCAGTCTGTGGCTCCAAAGGGGGGGCGCGGAGCCCTCAGAGGGGGGGCGG 68

QY 220 CGGCTGCTGCTTGGGCCC-GGCGAACTCCGGGGCCCTCTGTGGGAATGGGCCGGCTC 278
Db 69 CGGCTGCTGCTTGGGCCC-GGCGAACTCCGGGGCCCTCTGTGGGAATGGGCCGGCTC 128

QY 279 CGGGTCAAGGGGTGTAACCGCTCCCGGATGCGCGAGCGGCTTATCCGAGACCAAGTATGT 338
Db 129 CGGGTCAAGGGGTGTAACCGCTCCCGGATGCGCGAGCGGCTTATCCGAGACCAAGTATGT 188

QY 339 TGCCAGGACCGGAATGACACCTCAGGACCTTCCATGGGACCCCTCGCTATGGGGGA 398
Db 189 TGCCAGGACCGGAATGACACCTCAGGACCTTCCATGGGACCCCTCGCTATGGGGGA 248

QY 399 ACCCTTCAGTCCGACCTGGCTGGCCAGTCAGGATGGATCAGTCCCGCAAGACCTG 458
Db 249 ACCCTTCAGTCCGACCTGGCTGGCCAGTCAGGATGGATCAGTCCCGCAAGACCTG 308

QY 459 CCCCTCAGCAGATCCAGCAGGTCCAGCAGCGGTCCAAAATCGAAACCAATGCAA 518
Db 309 CCCCTCAGCAGATCCAGCAGGTCCAGCAGCGGTCCAAAATCGAAACCAATGCAA 368

QY 519 AGAAAAAGAGTGGCTGACAAAATTCACCTCAAAGGATTCGTGAATGGTACCAAGAT 578
Db 369 AGAAAAAGAGTGGCTGACAAAATTCACCTCAAAGGATTCGTGAATGGTACCAAGAT 428

QY 579 CCAGGCTATATGATCTCTGGCTTTGAAAGCAAACTGACACAGACTATCATGAGGA 638
Db 429 CCAGGCTATATGATCTCTGGCTTTGAAAGCAAACTGACACAGACTATCATGAGGA 488

QY 639 AACGGCTAGATATCAAGAGGCTTTGAAACGTCCCATTAAGCAAAACCGAAGCTGC 698
Db 489 AACGGCTAGATATCAAGAGGCTTTGAAACGTCCCATTAAGCAAAACCGAAGCTGC 548

QY 699 TTTTCATTTCAACTTTCAATCCGGCTAAGTCAAGTCCCGAGGATGGGAGGAGCGG 758
Db 549 TTTTCATTTCAACTTTCAATCCGGCTAAGTCAAGTCCCGAGGATGGGAGGAGCGG 608

QY 759 TGGCTTCCTGGGAGCTTCGGGTAGAGGACGGCTCCTGGAGGATTCAGCCTTGTCCAAAT 818
Db 609 TGGCTTCCTGGGAGCTTCGGGTAGAGGACGGCTCCTGGAGGATTCAGCCTTGTCCAAAT 668

QY 819 ATGATGCCCACTAAACAAAGAGGAAGTTCTCTTCC-TTCTTTAAGTCTCTGTGATTGAA 877

Db 669 ATGATGCCCACTAAACAAAGAGGAAGTTCTCTCTCTTTTAAAGTCTCTTGGTATTGAA 728
QY 878 CTGACAAAAGACCTGTATGGCCAGACAAACCATCTGTGTAGATGGCCAGACCGCCACT 937
Db 729 CTGACAAAAGACCTGTATGGCCAGACAAACCATCTGTGTAGATGGCCAGACCGCCACT 788
QY 938 ACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGATGTAATGTACGGTACT 997
Db 789 ACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGATGTAATGTACGGTACT 848
QY 998 GTCTTACTGATCTGGATTTACAGCCTCCCGAGTTTAAATTTAGACCCCGCTAGCTCGA 1057
Db 849 GTCTTACTGATCTGGATTTACAGCCTCCCGAGTTTAAATTTAGACCCCGCTAGCTCGA 908
QY 1058 TCCTTGGGCATCCATACCCAGACTCGTCCAGTATCATCAAGCACTGTGGCAATATATT 1117
Db 909 CTCCTGGGCATCCATACCCAGACTCGTCCAGTATCATCAAGCACTGTGGCAATATATT 968
QY 1118 AAGACACATAAGCTCCAGGACCCCTCAGAGCGGAGTTTGTCTATCTGTGACAACTACCTG 1177
Db 969 AAGACACATAAGCTCCAGGACCCCTCAGAGCGGAGTTTGTCTATCTGTGACAACTACCTG 1028
QY 1178 CAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCC 1237
Db 1029 CAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCC 1088
QY 1238 TTGCTTATGCCACAGAACCTATCATCTAATTAATCATCTCATCTGTGACCCCAATGAT 1297
Db 1089 TTGCTTATGCCACAGAACCTATCATCTAATTAATCATCTCATCTGTGACCCCAATGAT 1148
QY 1298 CAGAAAAAGACAGCTTGTATGACATTTGAACTGGATGACACCTTTGAAACCCAG 1357
Db 1149 CAGAAAAAGACAGCTTGTATGACATTTGAACTGGATGACACCTTTGAAACCCAG 1208
QY 1358 ATGAATTTCTTTCTGCTGCTCCACTGCCAGCCAAACAGGAGATTTGCTACTCTAGACAA 1413
Db 1209 ATGAATTTCTTTCTGCTGCTCCACTGCCAGCCAAACAGGAGATTTGCTACTCTAGACAA 1268
QY 1414 ----- 1413
Db 1269 ATCCATGAGACAAATAGAAACCATCAACAGCTGAAGACTCAGCGGGAGTTTCATGCTGAGC 1328
QY 1414 -----C 1414
Db 1329 TTTGCCAGAGACCTCAGGGTTTTCATCAATGACTGGCTTCAGTCCAGTCAGGAGACCTC 1388
QY 1415 AAGCAATGACTGATGTGGGGTAAACCCAGAGAGGAGCGCCAGCTGAGTTCTACTTTC 1474
Db 1389 AAGCAATGACTGATGTGGGGTAAACCCAGAGAGGAGCGCCAGCTGAGTTCTACTTTC 1448
QY 1475 CAGCCCTGGGCTCAGGAGGCTGTGTCCGATCTTCTACTCCAGAGTGCAGCAGAGAGA 1534
Db 1449 CAGCCCTGGGCTCAGGAGGCTGTGTCCGATCTTCTACTCCAAAGGTGCAGCAGAGAGA 1508
QY 1535 CAAGAAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGGCCCTCTCCACAGCCCTGAT 1594
Db 1509 CAAGAAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGGCCCTCTCCACAGCCCTGAT 1568
QY 1595 TCGACTGCACAAATTTCTTGATTTGGGCCCTGTGTGCTGCTGCCTCATAGTATCTGCCTTGG 1654
Db 1569 TCGACTGCACAAATTTCTTGATTTGGGCCCTGTGTGCTGCTGCCTCATAGTATCTGCCTTGG 1628
QY 1655 TCTTGTCTGGGGCTCCAGGGATGCTGTGGTTCAAGGACAGACAGCAATCAAGAGG 1714
Db 1629 TCTTGTCTGGGGCTCCAGGGATGCTGTGGTTCAAGGACAGACAGCAATCAAGAGG 1688
QY 1715 GTCTCACAAGACACCTGTTATCTCTTTCACCTTATCTCTTCCACCCAGCTTCC 1774
Db 1689 GTCTCACAAGACACCTGTTATCTCTTTCACCTTATCTCTTCCACCCAGCTTCC 1748
QY 1775 CTTTGGCCCAAAAGTTCCCATGTGCTGTACCTCCCTCGGTCTACATAGGACCTCTAG 1834

```

Db      1749  CTTTGGCCCAAAAGTTCCCATGTGCTGTACCCCTCCCTG3TCTACATAGGACCTCTAG 1808
      1835  ATAGTGTAGAGAGAGAACTAGTGTAAATAGTGTGAATGGATT-GGCCCTCAG 1893
      1809  ATATGTATAGAGAGAACTAGTGTATATAGTGTGATGATGATGGCCCTCAG 1868
      1894  CCAAGTGTCTTCMAAGGAGACAGTAACTGATCTTACCTTCAGAGACCCAGAGTTGG 1953
      1869  CCAAGTGTCTTCMAAGGAGACAGTAACTGATCTTACCTTCAGAGACCCAGAGTTGG 1928
      1954  G---TTTGCTCTCTCTCCAGACTCAGAGCTGTGGGCACTGTAAAGCTAGTTGATCTT 2010
      1929  GAGCTTCCCTCTCTCTCCAGACTCAGAGCTGTGGGCACTGTAAAGCTAGTTGATCTT 1988
      2011  GGCTCCCTGATTAAGAGATCCCAATTCCTCTCTCTCCAGAGTTGGAGAAACATC 2070
      1989  GGCTCCCTGATTAAGAGATCCCAATTCCTCTCTCTCCAGAGTTGGAGAAACATC 2048
      2071  TCCCTTCACTTGTGGCTGTAGCACTAGAGAAACCTGGTTCTT-GGCTCACTGAGCC 2129
      2049  TCCCTTCACTTGTGGCTGTAGCACTAGAGAAACCTGGTTCTTGGGCTCCACTGAGCC 2108
      2130  CCAAGTCACTCCAG-CCCTGTGGGTTGGCTGTAGTGTCTCTCACTCTCTAGT 2188
      2109  CCAAGTCACTCCAGCCCTGTGGGTTGGCTGTAGTGTCTCTCACTCTCTAGT 2168
      2189  TGGGTCACATCAGATTTGGAGTTTGTCTTATTTGCTCTCTCCAGACACTCCCTGT 2248
      2169  TGGGTCACATCAGATTTGGAGTTTGTCTTATTTGCTCTCTCCAGACACTCCCTGT 2228
      2249  GGCTGCCCTTGTGATTTCCCTCAGATCTGCCCTAATCCCGGAGATTGGGTGGGGAATC 2308
      2229  GGCTGCCCTTGTGATTTCCCTCAGATCTGCCCTAATCCCGGAGATTGGGTGGGGAATC 2288
      2309  TTGCCCTTCTCTTTCAGAGCCCGAGGATCTCATCTGGGGAACCTGTCACTTGCAGCAGAG 2368
      2289  TTGCCCTTCTCTTTCAGAGCCCGAGGATCTCATCTGGGGAACCTGTCACTTGCAGCAGAG 2348
      2369  GCTGTCTCTCTGCACTTTCAGAGATGTGATCTCATCTCACTCACTCCCTGCT 2428
      2349  GCTGTCTCTCTGCACTTTCAGAGATGTGATCTCATCTCACTCACTCCCTGCT 2407
      2429  CTGCATCCCTTAATGAGAAACGGGCTAAACCAACGGGTAACCAACGGGCTGAGCCAT 2488
      2408  CTGCATCCCTTAATGAGAAACGGGCTAAACCAACGGGTAACCAACGGGCTGAGCCAT 2466
      2489  CCTGTCTCTCTGCTCCCTTGTCTGCCCAGTTGACACTCTAGTGACTTCTAGGCACTG 2548
      2467  CCTGTCTCTCTGCTCCCTTGTCTGCCCAGTTGACACTCTAGTGACTTCTAGGCACTG 2526
      2549  AGGAGTGAAGGCTTAGGCTGTGAGATAGGCTGAGTTGGTTTGTGATCTTCTCTC 2608
      2527  AGGAGTGAAGGCTTAGGCTGTGAGATAGGCTGAGTTGGTTTGTGATCTTCTCTC 2586
      2609  TCCCTGCTCAGAGATTTGATCTCCAGGCTCCCTGAGCTCAAGGCTTGAAGCCCTGAG 2668
      2587  TCCCTGCTCAGAGATTTGATCTCCAGGCTCCCTGAGCTCAAGGCTTGAAGCCCTGAG 2646
      2669  GTAGCAGCAGAGCTTGTGATCTTGGCCCTTGTGATGATGATGATGATGATCTTCTC 2728
      2647  GTAGCAGCAGAGCTTGTGATCTTGGCCCTTGTGATGATGATGATGATGATCTTCTC 2706
      2729  AGGAGGCTCAGATTTCTTCTTCCAGGTTGATCACCCTCGAGTTGATATCCAGGCT 2788
      2707  AGGAGGCTCAGATTTCTTCTTCCAGGTTGATCACCCTCGAGTTGATATCCAGGCT 2766
      2789  CGCAGCTCAACAGCAGAGGCTGGAGCAGCTGGGCAACAAAGGGGATTC 2841
      2767  CGCAGCTCAACAGCAGAGGCTGGAGCAGCTGGGCAACAAAGGGGATTC 2819

```

RESULT 2
BQ933801

```

LOCUS      BQ933801                      929 bp    mRNA    linear    EST 21-AUG-2002
DEFINITION AGENCOURT 8823948 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423761
5' mRNA Sequence.
ACCESSION  BQ933801
VERSION    BQ933801.1 GI:22349184
KEYWORDS   EST.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 929)
AUTHORS   NIH-MGC http://mhc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: DCTD/DRP/Gazdar
          cDNA Library Preparation: Rubin Laboratory
          DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Place: LNC2602 row: n column: 18
          High quality sequence stop: 697.
FEATURES   Location/Qualifiers
            source          1..929
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:6423761"
                        /tissue_type="large cell carcinoma"
                        /lab_host="DH10B (phage-resistant)"
                        /clone_11b="NIH_MGC_18"
                        /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
                        EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                        into EcoRI/XhoI sites using the following 5' adaptor:
                        GGCAAGAG(G). Library constructed by Ling Hong in the
                        laboratory of Gerald M. Rubin (University of California,
                        Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                        Superscript II RT (Life Technologies). Note: this is a
                        NIH MGC Library."
BASE COUNT  205 a 260 c 293 g 170 t 1 others
ORIGIN
Query Match      31.4%; Score 892; DB 13; Length 929;
Best Local Similarity 99.1%; Pred. No. 5,9e-226;
Matches 917; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
      35  GCTTCCCTGCTGCGGCGCGCGCGGGAACAGCGGCGC-CGGGGGCGCTCGGGG 93
      1  GCTTCCCTGCTGCGGCGCGCGCGGGAACAGCGGCGCTCGGGGCGCTCGGGG 60
      94  GCGGGGAGGTTCCGGTTCCGGTTCTTGTGCGGCTGCATCGGCGGCTCCGGGAAGATG 153
      61  GCGGGGAGGTTCCGGTTCTTGTGCGGCTGCATCGGCGGCTCGGGGCGCTCGGGG 120
      154  CGGCGCGGCGGTTTCCAGTCTGTGCTCCAGCGGCGGCGCGGAGCTTCAGAGGAG 213
      121  CGGCGCGGCGGTTTCCAGTCTGTGCTCCAGCGGCGGCGCGGAGCTTCAGAGGAG 180
      214  CGGCGCGGCGGCTGCTGCTTGGGCGC-GGCGGAATCCGGGCGCTCTGTGGAATGGGCT 272
      181  CGGGGCGGCGGCTGCTGCTTGGGCGCGGGAACATCCGGGCGCTCTGTGGAATGGGCT 240
      273  CGGCTCCGGGCTCAAGGCTGTACCGCTCCCGGATCCCGGAGCGGCTATCGAGACAG 332
      241  CGGCTCCGGGCTCAAGGCTGTACCGCTCCCGGATCCCGGAGCGGCTATCGAGACAG 300
      333  GTATGTTCCAGGCGCGGAATGACACTTCAGGAGCTTTCATGAGGAGCCCTGCTATG 392
      301  GTATGTTCCAGGCGCGGAATGACACTTCAGGAGCTTTCATGAGGAGCCCTGCTATG 360

```

```
Qy 393 GGGGGAACCCCTTTCAGTCCGACCTCGCTGGCCAGCTCAGGAGTGGATCAGTCCCGCAAGA 452
Db 361 GGGGGAACCCCTTTCAGTCCGACCTGGCTGGCCAGTCAAGGATGGATCAGTCCCGCAAGA 420
Qy 453 GACCTGCCCCCTCAGCAGATCCAGCAGGTCCTCAGCAGCAGCGGTCCAAATCGAAACCAACA 512
Db 421 GACCTGCCCCCTCAGCAGATCCAGCAGGTCCTCAGCAGCAGCGGTCCAAATCGAAACCAACA 480
Qy 513 ATGCAAGAGAAAAGAGAGTGGCTGACAAAATTTCTACCTCAAGAGGATTCGTGAACCTGGTAC 572
Db 481 ATGCAAGAGAAAAGAGAGTGGCTGACAAAATTTCTACCTCAAGAGGATTCGTGAACCTGGTAC 540
Qy 573 CAGAAATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAACTGACACAGACTATCA 632
Db 541 CAGAAATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAACTGACACAGACTATCA 600
Qy 633 TGAGGAAACGGCTAGATATCCAGAGGCTTGAACGCTCCCATTTAAGCAAAAACGGAGAC 692
Db 601 TGAGGAAACGGCTAGATATCCAGAGGCTTGAACGCTCCCATTTAAGCAAAAACGGAGAC 660
Qy 693 TCGGAATTTTCAATTTCAACACTTTCAATCCGGCTAAGTCAAGTCCCGAGGATGGGAAG 752
Db 661 TCGGAATTTTCAATTTCAACACTTTCAATCCGGCTAAGTCAAGTCCCGAGGATGGGAAG 720
Qy 753 GGACGGTGGCTTCCCTGGAGCTTCGGGTAGAGGACGGCTCCTGAGGATTCAGCCTTGT 812
Db 721 GGACGGTGGCTTCCCTGGAGCTTCGGGTAGAGGACGGCTCCTGAGGATTCAGCCTTGT 780
Qy 813 CCAATATGATGCCACTAAACAAAGAGGAGTCTCTTCTTTTAAAGTCCCTTGGTGA 872
Db 781 CCAATATGATGCCACTAAACAAAGAGGAGTCTCTTCTTTTAAAGTCCCTTGGTGA 840
Qy 873 TTGAACCTGCAAAAGACCTGTATGGCCAGACCAACCATCTGGTAGAATGGCACAGGACCG 932
Db 841 TTGAACCTGCAAAAGACCTGTATGGCCAGACCAACCATCTGGTAGAATGGCACAGGACCG 900
Qy 933 CCACCTACCCAGGAGACCGATGGCTT 957
Db 901 CCACCTACCCAGGAGACCGATGGCTT 925

RESULT 3
LOCUS BX386559 933 bp mRNA linear EST 08-MAY-2003
DEFINITION BX386559 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CS0DJ005YB22 5-PRIME, mRNA sequence.
ACCESSION BX386559
VERSION BX386559.1 GI:30455451
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6875.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS1AJ002B120p1&cluster=6875.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS1AJ002B120p1.
FEATURES
Location/Qualifiers
1..933
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
/clone="CS0DJ005YB22"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 230 a 245 c 275 g 177 t 6 others
ORIGIN
Query Match 29.8%; Score 847.8; DB 13; Length 933;
Best Local Similarity 96.7%; Pred. No. 3.5e-214;
Matches 869; Conservative 6; Mismatches 23; Indels 1; Gaps 1;
Qy 112 CCGGTTCTTTGTGGCGGTGCATCGGCGCTCCGCGAAGATGGCGCGCGGGTTCCT 171
Db 35 CAGGCTGGTACCGGTCGGGAATTCCTGGGATCGGGAAGATGGCGCGCGGGTTCCT 94
Qy 172 AGTCTGTGGCTCCAAAGCGCGCGCGAGCCTCAGAGGGCGGGCGCGCTCGCT 231
Db 95 AGTCTGTGGCTCCAAAGCGCGCGCGAGCCTCAGAGGGCGGGCGCGCTCGCT 154
Qy 232 TGGGCC- GGCAGAACTCCCGGGGCTCTCTGTGCGAATGGGCTCCGGGTCAGAGGC 290
Db 155 TGGGCCCGGGCGGAACCTCCGGGGCTCTCTGTGCGAATGGGCTCCGGGTCAGAGGC 214
Qy 291 TGTAACGCTCCCGATCCCGGAGCGGCTATCCGAGACAGAGTATTTGCGAGGACGC 350
Db 215 TGTACCGCTCCCGATCCCGGAGCGGCTATCCRARACAGAGTATTTGCGAGGACGC 274
Qy 351 GAATGACACCTCAGGGACCTTCATGGGACCCCTGGCTATGGGGAAACCTTCAGTCC 410
Db 275 GAATGACACCTCAGGGACCTTCATGGGACCCCTGGCTATGGGGAAACCTTCAGTCC 334
Qy 411 GACCTGGCTGGCCAGTCAGGATGGATCAGTCCCGCAAGAGACCTTCGCCCTCAGCAGA 470
Db 335 GACCTGGCTGGCCAGTCAGGATGGATCAGTCCCGCAAGARACCTTCGCCCTCAGCAGA 394
Qy 471 TCCAGAGGTCAGCAGCAGCGGTCCAAATCGAAACCAATGCAAGAAAAAGAAAGA 530
Db 395 TCCAGAGGTCAGCAGCAGCGGTCCAAATCGAAACCAATGCAAGAAAAAGAAAGA 454
Qy 531 TGGCTGACAAAATTTCTACCTCAAGAGTTCGTGAACCTGGTACAGATCCAGCGCTATA 590
Db 455 TGGCTGACAAAATTTCTACCTCAAGAGTTCGTGAACCTGGTACAGATCCAGCGCTATA 514
Qy 591 TGGATCTCTTGGCTTTTGAAGGAAACTGGACCAAGACTATCATGAGGAAACGGCTAGATA 650
Db 515 TGGATCTCTTGGCTTTTGAAGGAAACTGGACCAAGACTATCATGAGGAAACGGCTAGATA 574
Qy 651 TCCAGAGGCTTGAAACGTCCTCCATTAAGCAAAAACGGAAGCTCGAATTTTCAATTTCTA 710
Db 575 TCCAGAGGCTTGAAACGTCCTCCATTAAGCAAAAACGGAAGCTCGAATTTTCAATTTCTA 634
Qy 711 ACCTTTTCAATCCGGCTAAGTCCAGATCCGAGGATGGGAAGGACCGTTCCTGGG 770
Db 635 ACCTTTTCAATCCGGCTAAGTCCAGATCCGAGGATGGGAAGGACCGTTCCTGGG 694
Qy 771 AGCTTCGGGTAGAGGACGGCTCCTCGAGGATTCAGGCTTTGTCCAAATATGATGCCACTA 830
Db 695 AGCTTCGGGTAGAGGACGGCTCCTCGAGGATTCAGGCTTTGTCCAAATATGATGCCACTA 754
Qy 831 AACAAAAAGAGAGTTCCTTCCTCTTTTAAAGTCCCTTTGATGATGAACGACCAAGACC 890
Db 755 AACAAAAAGAGAGTTCCTTCCTCTTTTAAAGTCCCTTTGATGATGAACGACCAAGACC 814
Qy 891 TGTATGGGCGAGACCAACCATCTGTAGAAATGGCACAGGACCGCCACTACCCAGGACCG 950
Db 815 TGTATGGGCGAGACCAACCATCTGTAGAAATGGCACAGGACCGCCACTACCCAGGACCG 874
```

RESULT 4	
LOCUS	BX375270
DEFINITION	BX375270 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CSODC011YE17 5-PRIME, mRNA sequence.
ACCESSION	BX375270
VERSION	BX375270.1 GI:30446410
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1201) Li W.B., Gruber C., Jesssee J. and Poljays D. Full-length cDNA libraries and normalization Unpublished
REFERENCE	Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr Library was constructed by life technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6875.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/ccluster.cgi?seq=CSODC011AC090Plc&cluster=6875.f . Contact : Peng Liang Email : liang@lifetech.com URL : http://fulllength.invitrogen.com/ Paraday Avenue Genoscope Sequence ID : CSODC011AC09QPI.
AUTHORS	Location/Organism
TITLE	1. 1201 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODC011YE17" /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED" /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
JOURNAL	
COMMENT	
BASE COUNT	308 a 297 c 309 g 227 t 60 others
ORIGIN	
Query Match	29.5%; Score 839.4; DB 13; Length 1201;
Best Local Similarity	98.0%; Pred. No. 6,5e-212;
Matches	887; Conservative 4; Mismatch 10; Indels 4; Gaps 4
Db	145 GGAAGATGGCGGCGCGGGGCTTCACAGTCTGTGGCTCCAGCGGGCGCGCGGAGCCT 204
Oy	69 GGAAGATGGCGGCGCGGGGCTTCACAGTCTGTGGCTCCAGCGGGCGCGCGGAGCCT 128
Oy	205 CAGAAGGGCGGCGCGGCTGTGCTGCTTGAGGCC-GGCGAACTCCGGGACTCTGTGC 263
Db	129 CAGAAGGGCGGCGGCTGTGCTGCTTGAGGCCGGGAGCACTCCGGGCTCTGTGC 188
Oy	264 GAATGGGCCCCGCTCCGGGTCAAAGGCTTACCCTCCCAGTAGCCCGAGGGGCTATC 323
Db	189 GAATGGGCCCCGCTCCGGGTCAAAGGCTTACCCTCCCAGTAGCCCGAGGGGCTATC 248
Oy	324 CGAGACCAAGTATGTTGCAGGAGCGCGAATACAACCTCAGGACCTTCATGGGACCCC 383
Db	249 CGAGACCAAGTATGTTGCAGGAGCGCGAATACAACCTCAGGACCTTCATGGGACCCC 308
Oy	384 CTGGCTATGGGGGAAAACCTTCAGTCCGACCTGGCTGGCCCAAGTACGGATGATCACT 443
Db	309 CTGGCTATGGGGGAAAACCTTCAGTCCGACCTGGCTGGCCCAAGTACGGATGATCACT 368
Oy	444 CCCGCAAGAACCTGCGCCCTCACAGCATCCAGCAGTCCAGACGACGAGCGGCTCCAATT 503

[illegible]

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6482661"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      255 a   231 c   223 g   198 t
ORIGIN
Query Match      29.3%; Score 833.6; DB 13; Length 907;
Best Local Similarity 97.6%; Pred. No. 2.1e-210;
Matches 867; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

Qy 439 TCAGTCCCGCAAGAGACCTGCCCTCAGCAGATCCAGCAGGTCCAGCAGCGCGGTCCA 498
Db 1 TCAGTCCCGCAAGAGACCTGCCCTCAGCAGATCCAGCAGGTCCAGCAGCGCGGTCCA 60

Qy 499 AAATCGAAACACATGCAAGAAAGAAAGAGATGGCTGACAAAATTCCTACCTCAAGGAT 558
Db 61 AAATCGAAACACATGCAAGAAAGAAAGAGATGGCTGACAAAATTCCTACCTCAAGGAT 120

Qy 559 TCGTGAACTGGTACCAGAAATCCAGCGCTATATGATCTCTTGGCTTTGAAAGGAAACT 618
Db 121 TCGTGAACTGGTACCAGAAATCCAGCGCTATATGATCTCTTGGCTTTGAAAGGAAACT 180

Qy 619 GCACAGACTATCATGAGAAACGGCTAGATATCCAGAGGCTTGAACGTCCTCATTA 678
Db 181 GCACAGACTATCATGAGAAACGGCTAGATATCCAGAGGCTTGAACGTCCTCATTA 240

Qy 679 GCAAAACGGAGCTGCGAAATTTTCATTTCTTAACACTTTTCAATCCGGTAACTAGTCAGATGC 738
Db 241 GCAAAACGGAGCTGCGAAATTTTCATTTCTTAACACTTTTCAATCCGGTAACTAGTCAGATGC 300

Qy 739 CGAGATGGGGAAGGAGCGGTGGCTTCTCGGAGCTTCGGGTAGAAAGACGGCTCCTGGA 798
Db 301 CGAGATGGGGAAGGAGCGGTGGCTTCTCGGAGCTTCGGGTAGAAAGACGGCTCCTGGA 360

Qy 799 GAATTCAGCTTGTCCAAATATGATGCGCACTAAACAAAGAGGAAGTTCTCTTCTTTT 858
Db 361 GAATTCAGCTTGTCCAAATATGATGCGCACTAAACAAAGAGGAAGTTCTCTTCTTTT 420

Qy 859 TAACTCTTGGTATGATGAGCAAGACCTGTATGGCCAGACACCACTATCGGTAGA 918
Db 421 TAACTCTTGGTATGATGAGCAAGACCTGTATGGCCAGACACCACTATCGGTAGA 480

Qy 919 ATGGCAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGA 978
Db 481 ATGGCAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGA 540

Qy 979 TGTGAATGTACGGTGTATGTCCTACTGATGCTGGAATACAGCCTCCCAAGTTTAAATT 1038
Db 541 TGTGAATGTACGGTGTATGTCCTACTGATGCTGGAATACAGCCTCCCAAGTTTAAATT 600

Qy 1039 AGACCCCGCTAGCTCGACTCCTGGGCATCCATACCAGACTCGTCAGGTGATCATCCA 1098
Db 601 AGACCCCGCTAGCTCGACTCCTGGGCATCCATACCAGACTCGTCAGGTGATCATCCA 660

Qy 1099 AGCACTGTGGCAATATATTAAGACACATAAGCTCCAGGACCCCTCACGA-GCGGAGTTTG 1157
Db 661 AGCACTGTGGCAATATATTAAGACACATAAGCTCCAGGACCCCTCACGA-GCGGAGTTTG 720

Qy 1158 TCATCTGTGACAAGTACTCGCAGCAGATCTTTGAGACTCAACGATATGAAGTTTTCAGAGA 1217
Db 721 TCATCTGTGACAAGTACTCGCAGCAGATCTTTGAGCTCCACGATATGAAGTTTTCAGAGA 780
```

```
Qy 1218 TCCCTCAGCGCTCCATGCTTGTATATGCACGAACTATCATCATTAATCAT-GTC 1276
Db 781 TCCCTCAGCGCTCCATGCTTGTATATGCACGAACTATCATCATTAATCATCTCTGGC 840

Qy 1277 ATCAGTGTTCACCGAATGATCAGAAAAAGACAGCTTTGTTATGACATT 1324
Db 841 ATCAGTGTTCACCGAATGATCAGAAAAAGACAGCTTTGTTATGACACT 888

RESULT 6
AL521992 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
LOCUS AL521992 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DB007YM01 5-PRIME, mRNA sequence.
ACCESSION AL521992
VERSION AL521992.2 GI:31040265
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 970)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12785485.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6875.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DB007AG01QPI&cluster=6875.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DB007AG01QPI.
FEATURES
Location/Qualifiers
source 1..970
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB007YM01"
/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 185 a 292 c 224 g 269 t
ORIGIN
```

```
Query Match      29.0%; Score 825.2; DB 9; Length 970;
Best Local Similarity 98.8%; Pred. No. 3.7e-208;
Matches 895; Conservative 0; Mismatches 3; Indels 8; Gaps 6;

Qy 1897 GGTGGTCTTCAAGGGACACAGCTAACTGATCCTTACCTTCAGAGACCAGAGTTGG-- 1954
Db 66 GATGGTCTTCAAGGGACACAGCTAACTGATCCTTACCTTCAGAGACCAGAGTTGGAG 125

Qy 1955 -TTTCGCTCTTCCAGAGCTCAGGCTGTGGGACTCTATAGCTAGTTGATCTTGGC 2013
Db 126 CTTTCGCTCTTCCAGAGCTCAGGCTGTGGGACTCTATAGCTAGTTGATCTTGGC 185

Qy 2014 TCTCCTGATAACAGATCCCAATTTCTTCTTCTTCTCCCTCCAGAGTTTGGAACTCTCC 2073
Db 186 TCTCCTGATAACAGATCCCAATTTCTTCTTCTTCTCCCTCCAGAGTTTGGAACTCTCC 245

Qy 2074 CTTCACTGTGTGCTGTAGCAGTACAGAAACCTCTGTTCTT-GGCTCCACTGAGCCCCA 2132
Db 246 CTTCACTGTGTGCTGTAGCAGTACAGAAACCTCTGTTCTTCTTGGGCTCCACTGAGCCCCA 305
```



```

661 CTGGCAATTTTCAATTTCAATCGGTAAGTCAGATGCCGAGGATGGGAA 720
QY 752 --GGGACGCTGCTCTGGGAG-CTTCGGGTAGAGACGCTCCTGGAGATTACGCC 808
Db 721 AGGGAGCGTGGCTTCTTGGGAGACTCCGGGTAAAGGACGCTCCTGGAAGATTACGCC 780
QY 809 TTGTCCTCAATATGATGCCCACTAAACAAAGAGGAGTTCCTCTCTCTTTTAAAGTTCCTTG 868
Db 781 TTGTCCTCAATATGATGCCCTTAAACAAAGAGGAGTTCCTCTCTTTTAAAGCCTTG 840
QY 869 GTGATTGAATGGACAAAGACTGTAT--GGGCCAGACAAACCATCTGTGTAGTAATGGCACAG 927
Db 841 GGGATTGGACCTGACCAAGACTGTATGGGCGCAACACCTTCTGTGTAAAGAGGACAG 900
QY 928 GACCGCCACTACCCAGGAGACCGATGG 954
Db 901 GAACCGCCTTACCCAGGACCCATGG 927

RESULT 8
BU845642
LOCUS
DEFINITION AGENCOURT_10414455 NIH_MGC_109 Homo sapiens cDNA clone
IMAGE: 6579312 5', mRNA sequence.
ACCESSION BU845642
VERSION BU845642.1 GI:24030083
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 953)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC#2782 row: 9 column: 24
High quality sequence stop: 661.
Location/Qualifiers
1..953
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6579312"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 227 a 260 c 282 g 178 t
ORIGIN
Query Match 28.5%; Score 809.2; DB 13; Length 953;
Best Local Similarity 98.0%; Pred No. 6.6e-204;
Matches 828; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
QY 135 GCGCGCTCCGGGAAGATCGCGCGCGGGTTCAGTCTGTGGCTCCAGCGCGGC 194

```

```

11 GGCGGCTCCGGGAAGATGGCGGCCCGGGCTTTTCAGTCTGTGGTCCAAAGCGGCGC 70
QY 195 GCGGAGCCTCAGAGGGGGCGGCGCGCTGCTGCTTGGGCC- GCGGAACTCCCGGG 253
Db 71 GCGGAGCCTCAGAGGGGGCGGCGCGCTGCTGCTTGGGCCGCGGGGNACTCCGGGG 130
QY 254 CTTCTGTGTGAATGGGCTCCGGCTCCGGGTCAAGGGGTGTACCGCTCCCGATCCCGGA 313
Db 131 CTTCTGTGTGAATGGGCTCCGGCTCCGGGTCAAGGGGTGTACCGCTCCCGATCCCGGA 190
QY 314 GCGGCTATCCGAGACCAAGTATGTTGCCAGGAGCGGATGACATCCAGGACCTTCC 373
Db 191 GCGGCTATCCGAGACCAAGTATGTTGCCAGGAGCGGATGACATCCAGGACCTTCC 250
QY 374 ATGGGACCCCTGGCTATGGGGGAAACCTTTCAGTCCGACTGCGCTGGGCCAGTCAAGG 433
Db 251 ATGGGACCCCTGGCTATGGGGGAAACCTTTCAGTCCGACTGCGCTGGGCCAGTCAAGG 310
QY 434 ATGGATCAGTCCCGCAAGACCTGCGCTCAGCAGATCCAGAGGTCAGAGCAGGCG 493
Db 311 ATGGATCAGTCCCGCAAGACCTGCGCTCAGCAGATCCAGAGGTCAGAGCAGGCG 370
QY 494 GTCCAAATCGAAACCAATGCAAGAAAAAGAGATGCTGACAAAAATTTACCTCAA 553
Db 371 GTCCAAATCGAAACCAATGCAAGAAAAAGAGATGCTGACAAAAATTTACCTCAA 430
QY 554 AGGATTCGTGAACCTGGTACCAGAAATCCAGGCGCTATATGATGATCTCTTGGCTTTTGAAGG 613
Db 431 AGGATTCGTGAACCTGGTACCAGAAATCCAGGCGCTATATGATGATCTCTTGGCTTTTGAAGG 490
QY 614 AAATCGACAGACTATCATGAGGAAACGGCTAGATATCAAGAGGCGCTTGAACCGTCC 673
Db 491 AAATCGACAGACTATCATGAGGAAACGGCTAGATATCAAGAGGCGCTTGAACCGTCC 550
QY 674 ATTAAGCAAAAACGGAAGCTGCGAATTTTCAATTTCAACACTTTCATTCGGCTTAAGTCA 733
Db 551 ATCAAGCAAAAACGGAAGCTGCGAATTTTCAATTTCAACACTTTCATTCGGCTTAAGTCA 610
QY 734 GATGCGGAGGATGGGAAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAGGACGGCTC 793
Db 611 GATGCGGAGGATGGGAAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAGGACGGCTC 670
QY 794 CTGGAGGATTCAGCCTTGTCCAAATATGATGCCCACTAAACAAAGAGGAAATTTCTTTC 853
Db 671 CTGGAGGATTCAGCCTTGTCCAAATATGATGCCCACTAAACAAAGAGGAAATTTCTTTC 730
QY 854 TTCTTTAAGTCTTGTGATTGAATGACAGCAAGACCTGTATGGGCCAGACCAATCTG 913
Db 731 TTTTAAAGTCTTGTGATTGAATGACAGCAAGACCTGTATGGGCCAGACCAATCTG 790
QY 914 GTAGAATGGCACAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGCCA 973
Db 791 GTAGAATGGCACAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGCCA 850
QY 974 GGAGA 978
Db 851 GGGGA 855

RESULT 9
AK075611
LOCUS
DEFINITION Mus musculus adult male brain cDNA, RIKEN full-length enriched
library, clone:071008A09 product:SWI/SNF related, matrix
associated, actin dependent regulator of chromatin, subfamily d,
member 1, full insert sequence.
ACCESSION AK075611
VERSION AK075611.1 GI:26344410
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```


REFERENCE	1	Carrinci, P. and Hayashizaki, Y.
AUTHORS		High-efficiency full-length cDNA cloning
TITLE		Meth. Enzymol. 303, 19-44 (1999)
JOURNAL		99279253
MEDLINE		10349636
PUBMED		
REFERENCE	2	Carrinci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
AUTHORS		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL		prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE		Genome Res. 10 (10), 1617-1630 (2000)
PUBMED		20499374
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carrinci, P.,
AUTHORS		Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
JOURNAL		Sun, N., Ishi, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A.,
MEDLINE		Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasinagi, K.,
PUBMED		Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,
REFERENCE	4	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
AUTHORS		Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL		sequencing pipeline with 384 multicapillary sequencer
MEDLINE		Genome Res. 10 (11), 1757-1771 (2000)
PUBMED		20530913
REFERENCE	5	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishi, Y.,
AUTHORS		Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, T., Fukuda, S.,
JOURNAL		Aizawa, K., Iwata, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I.,
MEDLINE		Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,
PUBMED		Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
REFERENCE	6	Flatschman, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H.,
AUTHORS		Kuehl, P., Lewis, S., Matsuo, Y., Nkaido, I., Pesole, G.,
JOURNAL		Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,
MEDLINE		Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
PUBMED		Baldarelli, R., Baren, G., Blake, J., Boffelli, D., Bojunga, N.,
REFERENCE	7	Carrinci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
AUTHORS		Flatschman, C., Fujita, M., Gariboldi, M., Gestrich, S., Hill, D.,
JOURNAL		Hotmann, M., Hume, D.A., Kamlay, M., Lee, N.H., Lyons, P.,
MEDLINE		Marchionni, L., Mashima, J., Mazzarelli, J., Mombereis, P., Nordone, P.,
PUBMED		Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
REFERENCE	8	Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
AUTHORS		Toyokawa, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
JOURNAL		Wysnar-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.
MEDLINE		and Hayashizaki, Y.
PUBMED		Functional annotation of a full-length mouse cDNA collection
REFERENCE	9	Nature 409 (6821), 685-690 (2001)
AUTHORS		21085660
JOURNAL		11217851
MEDLINE		
PUBMED		
REFERENCE	10	The FANTOM Consortium and the RIKEN Genome Exploration Research
AUTHORS		Group Phase I & II team.
TITLE		Analysis of the mouse transcriptome based on functional annotation
JOURNAL		of 60,770 full-length cDNAs
MEDLINE		Nature 420, 563-573 (2002)
PUBMED		6 (bases 1 to 2476)
REFERENCE	11	Adachi, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
AUTHORS		Arakawa, T., Bono, H., Carrinci, P., Fukuda, S., Fukunishi, Y.,
JOURNAL		Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
MEDLINE		Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
PUBMED		Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,
REFERENCE	12	Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
AUTHORS		Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Osa, C.,
JOURNAL		Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
MEDLINE		Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
PUBMED		Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
REFERENCE	13	Tejima, Y., Toyota, T., Yamamura, T., Yamana, I., Yasunishi, A.,
AUTHORS		Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (16-APR-2002) Yoshinide Hayashizaki, The Institute of
MEDLINE		Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT	
Exploration Research Group RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
Please visit our web site for further details.	
URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ URL:http://location.riken.go.jp/	
FEATURES	
source	1. .2476 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM DB:0710008A09" /db_xref="MGI:11900344" /db_xref="taxon:10090" /clone="0710008A09" /sex="male" /tissue_type="brain" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 1. .920 /note="unnamed protein product; SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1 (MGD MG1:1933623) putative" /codon_start=3 /protein_id="gAC35856.1" /db_xref="GI:26344411" /db_xref="MGI:1933623" /translation="VEGRILLEBALSKYDAITKOKRKSPFKSLVIELDQLYGPDNHN LVEHNRATTOETDGPQVYKRPGRDNNVACTULMDIYOPROPKDPRRLILGITQRR PVIIQALWOYIKTHKLDQPERHEFVLCDKTLQOI.FESQRMKFSSTFORHLALMPPEE IIINHVISVDNPDKCTACVDIDIVEDVDTLKTQNNPSFLSTAQOEIATDNKJLHETIT ETIQLTKTFEPMFSFARDPOGFINDMLOSCRDCLKTMDVGNPEEERAEFYQFOPM AOEAVCRFYFSKVQGRQRELEQALGINNT" 2463. 2468 polya_signal /note="putative" 2476 polya_site /note="putative" BASE COUNT 513 a 743 c 622 g 598 t ORIGIN
Query Match	28.0%; Score 795.6; DB 11; Length 2476;
Best Local Similarity	69.6%; Pred. No. 3.8e-200;
Matches 1530;	Conservative 0; Mismatches 409; Indels 259; Gaps 23;
QY	777 GGGTAGAAGCGGGCTCTGGAGGATTGACGCTTGCCAAATGATGAGCCACTAAACAA 836
DB	1 GGGTAGAAGCGGGATCCTGGAGGAACGGGCTTGCCAAATGATGAGCCCAAGCAAA 60
QY	837 AGAGGAATCTCTCTCTTTTAAGTCCTTGATGTGAATGGAGCAAGACCTGTATG 896
DB	61 AGAAGAATCTCTCTCTTTTAAGTCCTTGATGTGAATGGAGCAAGACCTGTATG 120
QY	897 GGGCAGACACCATCTGGTAGAATGGACACAGACCGCCACTACCCAGAGACCGATGGCT 956
DB	121 GCCCAGACACCATCTGGTAGAATGGACACAGACCGCCACTACCCAGAGACCGATGGCT 180
QY	957 TCCAGGTGAAGCGGCGAGAGATGTAATGATGAGCGTACTGTCTCTGATGCTGATTT 1016
DB	181 TCCAGGTGAAGCGGCGAGAGATGTAATGATGAGCGTACTGTCTCTGATGCTGATTT 240
QY	1017 ACCAGCTGCCCAATTTAAATTAGACCCCGCGCTAGCTGACTCTGGGATCCATACC 1076
DB	241 ACCAGCCCCCACAATTTAAATTAGACCCCGCGCTAGCTGACTCTGGGATCCATACC 300
QY	1077 AGACTCGTCACGATGATCCACGACTGTGGCACTATATTAAAGACACATTAAGCTCAGG 1136

Db 301 AGACACGTCAGATCATCAAGACCTGTGGCAGTATATTAACACACAAGCTCCAGG 360
Qy 1137 ACCCTCAGAGGGAGTTTCTCTGTGACAGTACCTGACAGATCTTTGAGATC 1196
Db 361 ACCCTCAGAGGGAGTTTCTCTGTGACAGTACCTGACAGATCTTTGATCTC 420
Qy 1197 AACGTATGAAGTTTTCAGAGATCCCTCAGCGCTCCATGCGCTTGTATGCCACAGAAC 1256
Db 421 AGCGATGAAGTTCTCAGAGATCCCTCAGCGCTCCATGCGCTTGTATGCCACAGAGC 480
Qy 1257 CTATCATATTAATCATGTATCATGTGTTGACCGATGATCAGAAAAGACAGCTTGT 1316
Db 481 CCATCATCATCAATCATGTATCATGTGTTGACCGATGATCAGAAAAGACAGCTTGT 540
Qy 1317 ATGACATTTGATTTGAAGTGGATGACCTTGAAGACCCAGATGAATCTTTCTGCTGT 1376
Db 541 ATGACATTTGATTTGAAGTGGATGACCTTGAAGACCCAGATGAATCTTTCTGCTGT 600
Qy 1377 CCATGCCAGGCCAACAGGAGATTGCTACTAGACAA 1413
Db 601 CCATGCCAGGCCAACAGGAGATTGCTACTAGACAA 660
Qy 1414 1413
Db 661 CCATCAACAGCTGAAGACCCAGAGAGTTTCTGAGCTTTGCCGAGACCTCAGG 720
Qy 1414 1413
Db 721 GTTTCATCAATGATGGTTTCACTCCAGTGGAGGACCTCAAGACGATGATGTTGG 780
Qy 1434 TGGGTAAACCCAGAGAGGAGCGCCAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGAGG 1493
Db 781 TGGGTAAACCCAGAGAGGAGCGCTGCTGAGTTCTACTTCCAGCCCTGGGCTCAGAGG 840
Qy 1494 TGTGTGCGGATCTTCTACTCAAGTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1553
Db 841 CTGTGTGCGGATCTTCTACTCAAGTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy 1554 TGGGAATCCGGAATACATAGGCGCTCCCAAGCGCTGATTCGAGTCCACCAATCTTG 1613
Db 901 TGGGAATCCGGAATACATAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 954
Qy 1614 ATTTGGGCGCTGTGCTGCTGCTCATAGTATCTGCTTGTCTTGTCTTGTGGGCGTCCA 1673
Db 955 -CTTGGGCGCTGTGCTGCTGCTCATAGTATCTGCTTGTCTTGTCTTGTGGGCGTCCA 1013
Qy 1674 GGGGATGCTGTGTTTCAAGGACAAAGACAGATGAAAGGGGTCTCA 1722
Db 1014 GGGGA---CTTGGGCTCAGGACAGTGTCAATGAAGAGGAGTGTCAATTTCTGTCTCA 1069
Qy 1723 -AGACACCTGTTATCT 1780
Db 1070 CAGTCACCTGTTATCCCGCTCTGTATCCCGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1129
Qy 1781 CCCAAGATTTCCCATGTGCTGTATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1840
Db 1130 ACCCACCCCGCTCAGCT 1189
Qy 1841 TTAGAGAGAGAAATGT 1900
Db 1190 ACATAGGACCTCTAGATAGTTAGAGAAACACAGAGTGGGGGCTCTCTGAGTCAAGTG 1249
Qy 1901 GTCTTCAAGGGACAGCTA-ACTATCTACCTT-CAGAGACCCAGAGTGGGTTTC 1958
Db 1250 GTCTTGGAGAGACAGCTATCACTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1309
Qy 1959 GCTCTCTCTCAAGACTCAGGCTGTGGGCTCTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2018
Db 1310 TATCTCTCTCAGCTCAGGCT-AGGCGAGTCTGTAAAGTAGTGTGACCTTGGGCTCTCC 1368
Qy 2019 TGATTAACAGAAATCAATTTCT 2078

Db 1369 CGGTAGC-----TTGGTCTCTTCCCTCCCTCCCGAGGTTGGG-----GCAGAGG 1413
Qy 2079 CTTGTTCCCTGTAGCATAAGAAACCTGGTTCTTGGGTCTCACTGAGCCCAAGGTGAG 2138
Db 1414 CTCTTTACCTCTGGCAGTAAAGAGC-----CTGGGCTTCACTGAGCCCGGTTGG 1466
Qy 2139 TCCCAAGCCTCTGGTGGGCTGCTGTCAGTGTCTCTCACTCTTATGTTGGGTGCAAC 2198
Db 1467 TCCCTCCCTCTGGACTTAACCTGCTGTCTAAGTGTCTCTGACCCCTTAGGGGTCCAT 1526
Qy 2199 ATCAGTATTCGAGTTTGTCT--TTATTGCTCCCTCCCGAGACACTCCCTGTGCTGCC 2256
Db 1527 GTCAGTATTCGAGTGTGTGTGAATTGTTGCTCCCTCCCGAGACTCCCGTAGCGGCCA 1586
Qy 2257 TTTGTGATTCCTCAGATCTGCCCTAATCCCGGCTTGGGTGGGGGAATCTTGGCTTT 2316
Db 1587 GTTGTAGATTTCCCTACACCTGCCCTTAACCCACGCTTTTGGGTGGG-ATCTTGCCTTT 1645
Qy 2317 CCCTTTCAGAGCCCGAGGATCTCATCTGGGAACTCTCATTTGCCAGAGGCTGTTC 2376
Db 1646 CC-----TTGTCTATTTCCAGCAGAGACTGTTC 1673
Qy 2377 TTCTGCACTTTGAGATGTGACTCATTTCCATTCACCTCACTCCACCTGCTCTGCACTC 2436
Db 1674 TTCTGCTGTAGAGGAGTGGCTT-----GTTTATTTCACTCCACCTGCCCTCTCCT 1725
Qy 2437 CTTTAAATGAGAAACCGGCTTAAACAAACCGGTTAAACAAAGCCCTGGGCTCTCTCT 2496
Db 1726 GTAAATGAGAAACAGGCTTGAATTAACCGGCT---AAAGCCCTAGGCCATCCCTGTCT 1782
Qy 2497 TCCTGCTCTTGTCTGCCAGTTGACACTTACTGTGTGACTTCTAGGGCACTGAGGAGTGA 2556
Db 1783 TCCTGCTCTGCTGCTGCCAGTTGAATCCCACTGGTGGCTTCCCGGCACTGAGGAGTAA 1842
Qy 2557 AAGCGCTAGGCTGAGAGATAGGCTGAGTGGGTTGTGACTCTTCTCCCTCTCCCTGCC 2616
Db 1843 AAGCGCTAGGCTGAGAGATAGGCTGAGTGGGTTGTGACTCTTCTCCCTCTCCCTGCC 1887
Qy 2617 TCACAGGATTTGACTCTCCCGAGCCCTCCCTCAAGCTTTCAGACCCCTCAGGTAGCAGC 2676
Db 1888 -----CCACCCCTGCTGCTCCCTCAAGCTTTCAGACCCCTCAGGTAGCAGC 1935
Qy 2677 AGGACCTTGTATCTTGGCCCTTGGATCTGAGATGTTTTTGCATCTTTCCAGGAGAGC 2736
Db 1936 AGGA-TGTGGGATCGAGGCCCTTGGGACAGA-----TGCTTTGAACTTTCCAGGGAAGC 1989
Qy 2737 CTACATCTTCTTCCAGTGTGTATCACCCCGAGTTAGCATATCCAGGCTCGCAGACT 2796
Db 1990 CTCGATTTCTC-----CAGGTTTGTACCCCGAGTTAGCATGTCCAGGCTCGCAGACA 2044
Qy 2797 CAACACAGCAAGGTTGGGAGAGAGCTGGGCAAAAGG 2834
Db 2045 ACAC---TGCAGGTTGGGAGACAGCTGGGCAACAGGGG 2079

RESULT 10
LOCUS B0881274
DEFINITION B0881274 972 bp mRNA linear EST 16-AUG-2002
ACCESSION AGENCOURT_8725685 NIH_MGC_47 Homo sapiens cdna clone IMAGE:6340339
VERSION 5', mRNA sequence.
KEYWORDS B0881274.1 GI:22273282
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 972)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L12M2536 row: j column: 20
 High quality sequence start: 36
 High quality sequence stop: 694.

FEATURES

source

1. 972
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6340339"
 /tissue_type="neuroblastoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 47"
 /note="Organ: Brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 262 a 246 c 240 g 223 t 1 others

Query Match 27.4% Score 779.8; DB 13; Length 972;
 Best Local Similarity 98.3%; Pred. No. 4.5e-196;
 Matches 798; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

608 GAAAGGAACTGGACCACTATCATGAGGAAACGGCTAGATATCCAGAGCCCTTGAA 667
 16 GAGGGGAACTGGACCACTATCCGTAAGAAACGGCTAGATATCCAGAGCCCTTGAA 75
 668 CGTCCCATTAAGCAAAAACGGAAGCTGCGAATTTTCACTTCAACCTTCAATCCGGCT 727
 76 CGTCCCATTAAGCAAAAACGGAAGCTGCGAATTTTCACTTCAACCTTCAATCCGGCT 135
 728 AATCTGATGCCAGATGAGGAGGAGGAGCGTGGCTTCTGGAGAGCTTGGGTAAGA 787
 136 AATCTGATGCCAGATGAGGAGGAGGAGCGTGGCTTCTGGAGAGCTTGGGTAAGA 195
 788 CGGCTCTGGAGATTCAGGCTTGTCCAAATATGATGCCACTAAACAAAAGAGAGTTC 847
 196 CGGCTCTGGAGATTCAGGCTTGTCCAAATATGATGCCACTAAACAAAAGAGAGTTC 255
 848 TCTTTC-TTCTTTAAGTCTTGTGATGATGAATGACAAAGACCTGTATGGCCAGACAA 906
 256 TCTTTC-TTCTTTAAGTCTTGTGATGATGAATGACAAAGACCTGTATGGCCAGACAA 315
 907 CCACTCTGATGATGACGACGACCGGCACTACCCAGAGAGCCGATGGCTTCCAGGTGA 966
 316 CCACTCTGATGATGACGACGACCGGCACTACCCAGAGAGCCGATGGCTTCCAGGTGA 375
 967 GCGGCGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026
 376 GCGGCGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435
 1027 CCAAGTTAAATTAGACCCCGGCTAGCTGATCTCTGGGATCCATACCGAGCTGTCC 1086
 436 CCAAGTTAAATTAGACCCCGGCTAGCTGATCTCTGGGATCCATACCGAGCTGTCC 495
 1087 AGGATCATCAAGACATGTCGCAATATTAAGACATTAAGACATTAAGACATTAAGACAT 1146
 496 AGGATCATCAAGACATGTCGCAATATTAAGACATTAAGACATTAAGACATTAAGACAT 555
 1147 GCGGAGTTTGTATCTGTGACAAAGTACCTGACGAGATTTTGAAGCTCAACGTATGA 1206

556 GCGGAGTTTGTATCTGTGACAAAGTACCTGACGAGATCTTGAAGCTCAACGTATGA 615
 1207 GTTTTCAGAGATCCCTCAGCGGCTCATGCTTGTATGACCAAGAACATCATCAT 1266
 616 GTTTTCAGAGATCCCTCAGCGGCTCATGCTTGTATGACCAAGAACATCATCAT 675
 1267 TAATCATGTCATCACTGTGACCCGAAATGATCAGAAAAGACAGCTTGTATGATGAT 1326
 676 TAATCATGTCATCACTGTGACCCGAAATGATCAGAAAAGACAGCTTGTATGATGAT 735
 1327 TGTGTAAGTGTGATGACACCTTAAAGACCAAGATGATTTTCTGCTCTCCACTGCC 1386
 736 TGTGTAAGTGTGATGACACCTTAAAGACCAAGATGATTTTCTGCTCTCCACTGCC 795
 1387 CCAACAGAGATGCTACTCTAGACACAGA 1418
 796 CCAACAGAGATGCTACTCTAGACACAGA 827

RESULT 11
 BU931457 872 bp mRNA linear EST 18-OCT-2002
 LOCUS AGENCOURT_10478213 NIH_MGC_109 Homo sapiens cDNA clone
 DEFINITION IMAGE:6671629 5', mRNA sequence.
 BU931457
 ACCESSION BU931457.1 GI:24120276
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 872)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L12M2950 row: n column: 13
 High quality sequence stop: 646.

FEATURES

source

1. 872
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6671629"
 /tissue_type="leukocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 109"
 /note="Organ: ovary; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 209 a 238 c 261 g 164 t

Query Match 27.2% Score 772.4; DB 13; Length 872;
 Best Local Similarity 98.7%; Pred. No. 4.1e-194;
 Matches 831; Conservative 0; Mismatches 6; Indels 5; Gaps 5;

139 GCTCCGGAAGATGCGCGCGCGGCTTTCAGTCTGTGCTTCAAGCGGCGCGCG 198
 1 GCTCCGGAAGATGCGCGCGCGGCTTTCAGTCTGTGCTTCAAGCGGCGCGCGCG 60

JOURNAL
COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLAM12346 row: e column: 02
High quality sequence stop: 647.

FEATURES
SOURCE

```
1. 1003
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TMACG:5583337"
/tissue_type="embryonal"
/lab_host="PH10 (phage)"
/clone_lib="NIH MGC 92"
/notes="Organ: testis; Vector:
Site_2: SalI; Cloned unit:
Average insert size 2.5
full-length clones and cDNA
Note: this is a NIH MGC
```

BASE COUNT	190 a	308 c	236 g	269 t
ORIGIN				

Query Match	27.2%;	Score 771.8;	DB 12;
Best Local Similarity	96.8%;	Pred. No. 6.1e-194;	
Marches 874.	Conservative	0.	Mismatches 17.
			Indels 12.
			Calls 8.

Qy	1490	GAGCCTGTGTGCCGATACCTTCTACTCAAGGTGCAGCAGAGACGACAGAAATTTAGACAA	1549
Db	1	GAGCCTGTGTGCCGATACCTTCTACTCAAGGTGCAGCAGAGACGACAGAAATTTAGACAA	60
Qy	1550	GCCTCGGAATCGGAAATACATAGGGCCTCTCCACAGACCCCTGATTTCGACTGCACCAAT	1609
Db	61	GCCTCGGAATCGGAAATACATAGGGCCTCTCCACAGACCCCTGATTTCGACTGCACCAAT	120
Qy	1610	CTTGATTTGGGCCCTGTGCTGCCTCATAGTATCTGCCTTTGGTCTTCTTTGGGGCGT	1669
Db	121	CTTGATTTGGGCCCTGTGCTGCCTCATAGTATCTGCCTTTGGTCTTCTTTGGGGCGT	180
Qy	1670	TCCAGGGATGCTGTTGGTTCAAGGACAGACAGAAATGAAGAGGTCTCACAAAGACACC	1729
Db	181	TCCAGGGATGCTGTTGGTTCAAGGACAAACACAGAAATGAAGAGGTCTCACAAAGACACC	240
Qy	1730	TGTTATCCTCTCTTTTCACCCCTATCTTCCACCCCGCCAGCTTCCCTTTGCCCCACAAAG	1789
Db	241	TGTTATCCTCTCTTTTCACCCCTATCTTCCACCCCGCCAGCTTCCCTTTGCCCCACAAAG	300
Qy	1790	TTCCCATGTGCCTGTACCCCTCCCTGGCTACATAGGACCTCTAGATAGTGTTTAGAGAGA	1849
Db	301	TTCCCATGTGCCTGTACCCCTCCCTGGCTACATAGGACCTCTAGATAGTGTTTAGAGAGA	360
Qy	1850	GACATGTAGTGGTAATAGTGTCTGGAAATGGATT--GGCCTCAGGCCAGGTGGTCTTCAA	1908
Db	361	GAACATGTAGTGGTAATAGTGTCTGGAAATGGAAATGGGCCCTCAGGCCAGGTGGTCTTCAA	420
Qy	1909	GGGACACAGCTAACTGATCCTACCCCTTCAGAGACCCAGGAGTTGGG--TTTTCGCTCCTT	1965
Db	421	GGGACACAGCTAACTGATCCTACCCCTTCAGAGACCCAGGAGTTGGGAGCTTTTCGCTCCTT	480
Qy	1966	CTCCAAGACTCAGGCCCTGTGGGCACCTATAAGCTAGTTGATCTTGGCTCTCCTGATAAC	2025
Db	481	CTCCAAGACTCAGGCCCTGTGGGCACCTATAAGCTAGTTGATCTTGGCTCTCCTGATAAC	540
Qy	2026	AGAAATCAAATTTCTTCTCCTTCCCTCCACAGGTTTGGAAACAACTCTCCCTTCACTTGTG	2085
Db	541	AGAAATCAAATTTCTTCTCCTTCCCTCCACAGGTTTGGAAACAACTCTCCCTTCACTTGTG	600

Qy	199	GAGCCTCAGAGAGGGCGGGCTGCTGCTTGGGCCCGCGGAACTTCGGGGCGCTC	255
Db	61	GAGCCTCAGAGAGGGCGGGCTGCTGCTTGGGCCCGGGGAACTTCGGGGCGCTC	120
Qy	258	CTGTGCGAATGGGCCCGGCTCCGGGTCAAGGGCTGTACCGTCCCGCATGCCGAGACGG	317
Db	121	CTGTGCGAATGGGCCCGGCTCCGGGTCAAGGGCTGTACCGTCCCGCATGCCGAGACGG	180
Qy	318	CTTATCCGAGACACAGGTATGTTGCCAGGGAGCCGAAATGACACTCAGGGACCTTCCATGG	377
Db	181	CTTATCCGAGACACAGGTATGTTGCCAGGGAGCCGAAATGACACTCAGGGACCTTCCATGG	240
Qy	378	GACCCCTCGCTATGGGGGAACCCCTCAGTCCGACCTGGCTGGGCCCGTCAAGGATGG	437
Db	241	GACCCCTCGCTATGGGGGAACCCCTCAGTCCGACCTGGCTGGGCCCGTCAAGGATGG	300
Qy	438	ATCAGTCCCGCAGAGACCTGCCCTCTCAGCAGATCCAGCAGGTCCAGCAGCGCGTCC	497
Db	301	ATCAGTCCCGCAGAGACCTGCCCTCTCAGCAGATCCAGCAGGTCCAGCAGCGCGTCC	360
Qy	498	AAAATCGAAACACAAATGCAAGAAAAAGATGGGTGACAAAATTTCTACCTCAAAGGA	557
Db	361	AAAATCGAAACACAAATGCAAGAAAAAGATGGGTGACAAAATTTCTACCTCAAAGGA	420
Qy	558	TTTCGTGAACTGGTACCAGAAATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAAC	617
Db	421	TTTCGTGAACTGGTACCAGAAATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAAC	480
Qy	618	TGGACCAGACTATCATGAGGAAACGGCTAGATATCCAAGAGGCTTGAACCGTCCCATTA	677
Db	481	TGGACCAGACTATCATGAGGAAACGGCTAGATATCCAAGAGGCTTGAACCGTCCCATTA	540
Qy	678	AGCAAAAAAGGAAGCTGCGAATTTTCATTTCTAACACTTTCAATCCGGCTTAAGTCAATG	737
Db	541	AGCAAAAAAGGAAGCTGCGAATTTTCATTTCTAACACTTTCAATCCGGCTTAAGTCAATG	600
Qy	738	CCGAGATGGGAAAGGACGGTGGCTCTCTGGGAGCTTCGGGTAGAGGACGGCTCCTGG	797
Db	601	CCGAGATGGGAAAGGACGGTGGCTCTCTGGGAGCTTCGGGTAGAGGACGGCTCCTGG	660
Qy	798	AGATTTCAGCTTGTCCAAATATGATGCCATAAAAGAGAGGAGTTCTCTTCC-TTC	856
Db	661	AGATTTCAGCTTGTCCAAATATGATGCCATAAAAGAGAGGAGTTCTCTTCC-TTC	720
Qy	857	TTTAAAGTCTTGGTGAATTTGAACTGGACAAAGACCTGTATGGGCCAGACACCATCTGTA	916
Db	721	TTTAAAGTCTTGGTGAATTTGAACTGGACAAAGACCTGTATGGGCCAGACACCATCTGTA	780
Qy	917	GAATGGCACAGGA-CGCGCACCTACCC-AGGAGACCGAT-GGCTTCCAGGTGAAGCGGCCA	973
Db	781	GAATGGCACAGGACCGCCACTACCAAGAGGAGCCGATGGGCTTTCAGGTGAAGCGGCC	840
Qy	974	GG 975	
Db	841	GG 842	

RESULT 12	
BM457092	
LOCUS	1003 bp mRNA linear EST 05-FEB-2002
DEFINITION	AGENCOCURT 6406599 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583337
	5', mRNA sequence.
ACCESSION	BM457092
VERSION	BM457092.1 GI:18506132
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 1003)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)

Db	301	TTGGATCAGTCCCGAAGAGAGACCTGCCCTCAGCAGATCCAGCAGGTCCAGCAGCAGGCGG	360
Qy	495	TCCAAAATCGAAAACCAATGCAAGAAAAAGAGATGGCTGACAAAAATTTACTCTCAA	554
Db	361	TCCAAAATCGAAACCAATGCAAGAAAAAGAGATGGCTGACAAAAATTTACTCTCAA	420
Qy	555	GGATTTCGTGAACCTGGTACCAAGATCCAGAGCTATATGGATCTCTTGGCTTTTGAAGGA	614
Db	421	GGATTTCGTGAACCTGGTACCAAGATCCAGAGCTATATGGATCTCTTGGCTTTTGAAGGA	480
Qy	615	AATCGACACAGACTATCATGAGGAAACGGCTAGATATCCAAGAGCCTTGAAGCGTCCCA	674
Db	481	AATCGACACAGACTATCATGAGGAAACGGCTAGATATCCAAGAGCCTTGAAGCGTCCCA	540
Qy	675	TTAAGCAAAAACGGAAGCTCGAATTTTCATTTCTAAACATTTTCAATCCGGCTAAGTCAG	734
Db	541	TCAAGCAAAAACGGAAGCTCGAATTTTCATTTCTAAACATTTTCAATCCGGCTAAGTCAG	600
Qy	735	ATCCGAGGATGGGAAAGGACGGTGGCTCTCTGGGAGCTTCGGTGAAGAGCGGTCC	794
Db	601	ATCCGAGGATGGGAAAGGACGGTGGCTCTCTGGGAGCTTCGGTGAAGAGCGGTCC	660
Qy	795	TGGAGGATTCAGCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGGAAGTTCTCTTCT	854
Db	661	TGGAGGATTCAGCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGGAAGTTCTCTTCT	720
Qy	855	TCCTTAAAGTCCTGGTGATTTGAACCTGGACAAAAGACCTGTATGGGCCAGACAAACCAT	912
Db	721	TTTCTTAAAGTCCTGGTGATTTGAACCTGGACAAAAGACCTGTATGGGCCAGACAAACCAT	780
Qy	913	GGTAGAATGGCACAGGACCGCCAC	936
Db	781	GGTAGAATGGCACAGGACCGCCAC	804
RESULT 15			
BQ949766			
LOCUS			
DEFINITION	BQ949766	878 bp mRNA linear EST 21-AUG-2005	
ACCESSION	AGENCOURT 8794796	NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6375035	
VERSION	BQ949766	5', mRNA sequence.	
KEYWORDS	BQ949766.1	GI:22365244	
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	1 (bases 1 to 878)		
COMMENT	NIH-MGC http://mgi.nci.nih.gov/.		
	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: DCTD/BTP/Gazdar		
	CDNA Library Preparation: Rubin Laboratory		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: L1CM2554 row: p column: 12		
	High quality sequence stop: 693.		
FEATURES	Location/Qualifiers		
source	1..878		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6375035"		
	/tissue_type="large cell carcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_18"		
	/note="Organ: lung; Vector: pOTB7; Site:1; XhoI; Site:2;		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 08:23:36 : Search time 4251 Seconds
(without alignments)
16590.966 Million cell updates/sec

Title: U66619
Perfect score: 1724
Sequence: 1 GAATTCGCGCGAGCGGCC.....GTCAATTCATCTGGAAATTC 1724

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hhg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hhg_hum:*
31: em_hhg_inv:*
32: em_hhg_other:*
33: em_hhg_mus:*
34: em_hhg_pln:*
35: em_hhg_rtd:*
36: em_hhg_mam:*
37: em_hhg_vrt:*
38: em_sy:*
39: em_hhgo_hum:*
40: em_hhgo_mus:*
41: em_hhgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1724	100.0	1724	9	HSU66619	U66619 Human SWI/S
2	1585.2	91.9	1698	9	BC002628	BC002628 Homo sapi
3	1365.8	79.2	1413	12	BT007694	BT007694 Synthetic
4	1198.6	69.5	1790	10	BC013122	BC013122 Mus muscu
5	701.2	40.7	3086	10	BC026783	BC026783 Mus muscu
6	690.6	40.1	2746	9	AF109733	AF109733 Homo sapi
7	690.6	40.1	3335	9	BC009368	BC009368 Homo sapi
8	685.2	39.7	3103	6	AX305314	AX305314 Sequence
9	685.2	39.7	3103	10	MMU66620	U66620 Mus musculu
10	679	39.4	2747	5	BC049347	BC049347 Danio rer
11	662.8	38.4	3101	10	MUSDI5K2L	M25773 Mus musculu
12	655.4	38.0	2297	10	AB003504	AB003504 Rattus no
13	649	37.6	2379	10	BC005732	BC005732 Mus muscu
14	634.4	36.8	2041	9	HSU66618	U66618 Human SWI/S
15	630.6	36.6	2557	5	BC045009	BC045009 Xenopus I
16	572.6	33.2	2026	3	AY069383	AY069383 Drosophi1
17	571	33.1	2018	3	AF071503	AF071503 Drosophi1
18	522.8	30.3	2841	9	HSU66617	U66617 Human SWI/S
19	504.2	28.2	33444	2	AC014246	AC014246 Drosophi1
20	504.2	28.2	171751	3	AC023697	AC023697 Drosophi1
21	504.2	29.2	174157	3	AC023711	AC023711 Drosophi1
22	504.2	29.2	332029	3	AE003491	AE003491 Drosophi1
23	413.2	24.0	137956	2	AC010918	AC010918 Drosophi1
24	410.4	23.8	450	6	AR240523	AR240523 Sequence
25	405.6	22.5	1674	3	AK112677	AK112677 Cioma int
26	276	16.0	1350	9	AK025917	AK025917 Homo sapi
27	235.2	13.6	247	11	G30760	G30760 SWS3165 Ex
28	234.8	13.6	44171	3	AF000265	AF000265 Caenorhab
29	234.8	13.6	299719	2	AC006858	AC006858 Caenorhab
30	216.5	12.6	294136	2	AC006901	AC006901 Caenorhab
31	179	10.4	157075	9	AC005486	AC005486 Homo sapi
32	174.2	10.1	214809	2	AC132744	AC132744 Rattus no
33	174.2	10.1	220259	2	AC134763	AC134763 Rattus no
34	174.2	10.1	263052	2	AC099360	AC099360 Rattus no
35	170.4	9.9	35899	9	AC021097	AC021097 Homo sapi
36	170.4	9.9	213045	2	AC093640	AC093640 Homo sapi
37	167.8	9.7	27832	3	CEZK1128	Z47357 Caenorhabd
38	167.8	9.7	239244	2	AC116136	AC116136 Mus muscu
39	140.8	8.2	1216	17	AF113019	AF113019 Homo sapi
40	137.8	8.0	310	6	AX098486	AX098486 Sequence
41	136.8	7.9	160431	8	OSJN00176	AL662977 Oryza sat
42	133	7.7	1121	9	BC018853	BC018853 Homo sapi
43	119.4	6.9	1898	8	AK119040	AK119040 Arabidops
44	117.6	6.8	187859	2	BX323804	BX323804 Danio rer
45	114.2	6.6	197095	2	BX322787	BX322787 Danio rer

ALIGNMENTS

RESULT 1
HSU66619
LOCUS
DEFINITION Human SWI/SNF complex 60 kDa subunit (BAF60c) mRNA, complete cds.
ACCESSION U66619.1 GI:1549246
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1724)
Wang, W., Cole, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R.,
Muehardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L. and
Crabtree, G.R.

Pred. No. is the number of results predicted by chance to have a

TITLE Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
JOURNAL EMBO J 15 (1996) In press
REFERENCE 2 (bases 1 to 1724)
AUTHORS Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R. and Crabtree, G.R.
TITLE Diversity and specialization of mammalian SWI/SNF complexes
JOURNAL Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE 96397413
PUBMED 8804307
REFERENCE 3 (bases 1 to 1724)
AUTHORS Wang, W., Xue, Y., Zhou, S. and Crabtree, G.R.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
FEATURES
source
1..1724
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="Jurkat T cells"
1..1724
/gene="BAF60c"
181..1590
/gene="BAF60c"
/note="Unlike BAF60a and BAF60b this gene is highly expressed in muscle cells; similar to human BAF60a, BAF60b, BAF60c and yeast SWP73 and SWP73b"
/codon_start=1
/product="SWI/SNF complex 60 kDa subunit"
/protein_id="AAC50697.1"
/db_xref="GI:1549247"
/translation="MTPGLQHPPTVVRPGMSPGARMHQGRPMAPRAPRTWAAPPDCD
PAMPEDGARPOASAPARQSOQOPEPTAPARSAPSRKMRKWLKILPQRIRELV
PESQAYMELLAFKLDQTIMRGVDIEALKRPKMKRKLRLYISNTFPAKSDAED
SDGSIASWELRVEGLKDDPSKRRKFSFFKSLVIELDKOLYEDNHLVHWKRPPT
QETDGFVKRKEIDSVRCTLMLLDYQPPFKLDPRLLRLGLHTQSRSAIVQALWYD
VKTNRLODSHPKIDYQKQIIFDCPRKLFSEIPORLALLLPDPPIVNHVSD
PSDOKTACYDIDIVEVEPLKQMSFLLSTANQOEISPLDSKIHTETESINQLKIQ
DFMLSFSDPKGYVDLRSQSLDKVMTDVAGNPEERRAEFYHQPWSQAVSRFYF
CKIQRRQELQSLVVRNT"
BASE COUNT 376 a 546 c 510 g 292 t
ORIGIN
Query Match 100.0%; Score 1724; DB 9; Length 1724;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1724; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAATTCGGCGCAGCGCCGAGCCGAGCGCCGAGCGAGCGCGCGCGCGCGCGCGCGCG 60
Db |||||
Qy 1 GAATTCGGCGCAGCGCCGAGCCGAGCGCCGAGCGAGCGCGCGCGCGCGCGCGCGCG 60
Db |||||
Qy 61 GCGGGGTCCGGGGGAGCAGATCTCAGATGGCCCTTGGTGTGCGAGGCGCGTGGGC 120
Db |||||
Qy 61 GCGGGGTCCGGGGGAGCAGATCTCAGATGGCCCTTGGTGTGCGAGGCGCGTGGGC 120
Db |||||
Qy 121 TCCGGGCCAGGCACCGAGGGGGGACTGGATGACTCTCCAGGTGAGGACCCCTGCCATCT 180
Db |||||
Qy 121 TCCGGGCCAGGCACCGAGGGGGGACTGGATGACTCTCCAGGTGAGGACCCCTGCCATCT 180
Db |||||
Qy 181 ATGACTCCAGGTCTTCAGCACCCACCCCGGTGTACAGCGCCCGGGATGCCGTCTGGA 240
Db |||||
Qy 181 ATGACTCCAGGTCTTCAGCACCCACCCCGGTGTACAGCGCCCGGGATGCCGTCTGGA 240
Db |||||
Qy 241 GCCCGATGCCACAGGGGGGCCATGGGCCCCCGGGTCCCGGTACATGGGCAGCC 300
Db |||||
Qy 241 GCCCGATGCCACAGGGGGGCCATGGGCCCCCGGGTCCCGGTACATGGGCAGCC 300
Db |||||
Qy 301 CCGCGTGGACCCCGGCTGGCCCCCGGGATGGAGCCCGCCCGCGCAAGCAGCAGCGCC 360
Db |||||
Qy 301 CCGCGTGGACCCCGGCTGGCCCCCGGGATGGAGCCCGCCCGCGCAAGCAGCAGCGCC 360
Db |||||
Qy 361 CCCGCCCGCAGAGCCAGCAAGAGCCAGGGCCAGCGGAGCCACCGCCCCCGCGCG 420
Db |||||

Db 361 CCCGCCCGCAGAGCCAGGCGCACAGAGCCAGGGCCAGCCGGAGCCACCGCCCCCGCGCG 420
Qy 421 AGCCCGCAGTGCACAGAGAGGAGGAAGATGGCTGACAAAATCTCTCCCTCAAAGGATTCCGGAG 480
Db |||||
Db 421 AGCCCGCAGTGCACAGAGAGGAGGAAGATGGCTGACAAAATCTCTCCCTCAAAGGATTCCGGAG 480
Qy 481 CTGCTCCCGAGTCCCAGGCTTACATGGACCTCTTGGCACTTGGAGGAAACTGGATCAA 540
Db |||||
Db 481 CTGCTCCCGAGTCCCAGGCTTACATGGACCTCTTGGCACTTGGAGGAAACTGGATCAA 540
Qy 541 ACCATCATGCGGAAGGGGTGGACATCCAGAGGCTCTGAAGAGCCCATGAAGCAAAAG 600
Db |||||
Db 541 ACCATCATGCGGAAGGGGTGGACATCCAGAGGCTCTGAAGAGCCCATGAAGCAAAAG 600
Qy 601 CGGAAGCTGCGACCTATATCTCCAAACACTTTTAACTGCGAAGTCTGATGCTGAGGAT 660
Db |||||
Db 601 CGGAAGCTGCGACCTATATCTCCAAACACTTTTAACTGCGAAGTCTGATGCTGAGGAT 660
Qy 661 TCCGACCGGACATTTGCTCTCTGGAGCTACGGGTGAGGGGAAGCTCTGGATGATGCC 720
Db |||||
Db 661 TCCGACCGGACATTTGCTCTCTGGAGCTACGGGTGAGGGGAAGCTCTGGATGATGCC 720
Qy 721 AGCAAAACAGAGCGGAAGTCTCTCTTTCTTCAAGAGTTTGGTCATCGAGCTGGACAAA 780
Db |||||
Db 721 AGCAAAACAGAGCGGAAGTCTCTCTTTCTTCAAGAGTTTGGTCATCGAGCTGGACAAA 780
Qy 781 GATCTTTATGGCCCTGACCAACCACTCGTTGAGTGGCATCGGACACCCACACCCAGGAG 840
Db |||||
Db 781 GATCTTTATGGCCCTGACCAACCACTCGTTGAGTGGCATCGGACACCCACACCCAGGAG 840
Qy 841 ACAGACCGCTTCCAGGTGAAAACCGCTGGGACCTGAGTGTGCGCTGCAACGCTGCTCCTC 900
Db |||||
Db 841 ACAGACCGCTTCCAGGTGAAAACCGCTGGGACCTGAGTGTGCGCTGCAACGCTGCTCCTC 900
Qy 901 ATGCTGAGTACACAGCCTCCAGTTCAAACCTGATCCCGCCTAGCCCGCTGCTGGG 960
Db |||||
Db 901 ATGCTGAGTACACAGCCTCCAGTTCAAACCTGATCCCGCCTAGCCCGCTGCTGGG 960
Qy 961 CTGCACACACAGAGCGCTCAGCCATTGTCTCAGGCGCTGTGGCAGTGTGTAAGACCAAC 1020
Db |||||
Db 961 CTGCACACACAGAGCGCTCAGCCATTGTCTCAGGCGCTGTGGCAGTGTGTAAGACCAAC 1020
Qy 1021 AGGCTGACGAGTCCCATGACAAAGGAATACATCAATGGGGAAGATTTTCCAGCAGATT 1080
Db |||||
Db 1021 AGGCTGACGAGTCCCATGACAAAGGAATACATCAATGGGGAAGATTTTCCAGCAGATT 1080
Qy 1081 TTTGATTTGTCGGCTGAAATTTTCTGAGATTTCCAGCGCCTCAGAGCCCTGCTATTG 1140
Db |||||
Db 1081 TTTGATTTGTCGGCTGAAATTTTCTGAGATTTCCAGCGCCTCAGAGCCCTGCTATTG 1140
Qy 1141 CCCCTGACCCAAATTTGTCATCAACCATGTCTCAGCGTGGACCCCTTACAGACAGAAAG 1200
Db |||||
Db 1141 CCCCTGACCCAAATTTGTCATCAACCATGTCTCAGCGTGGACCCCTTACAGACAGAAAG 1200
Qy 1201 ACAGCGTCTATGACATTTGAGTGGAGTGGAGGAGCCATTAAAGGGCAGATGAGCAGC 1260
Db |||||
Db 1201 ACAGCGTCTATGACATTTGAGTGGAGTGGAGGAGCCATTAAAGGGCAGATGAGCAGC 1260
Qy 1261 TTCTCTCTATCCAGGCCAACACAGAGGAGATCAGTCTCTGGACAGTAAAGTCCATGAG 1320
Db |||||
Db 1261 TTCTCTCTATCCAGGCCAACACAGAGGAGATCAGTCTCTGGACAGTAAAGTCCATGAG 1320
Qy 1321 ACAGTTGAGTCCATAAACACAGTCAAGATCAGAGGAGTCTTCTGCTAAAGCTTCTCCAGA 1380
Db |||||
Db 1321 ACAGTTGAGTCCATAAACACAGTCAAGATCAGAGGAGTCTTCTGCTAAAGCTTCTCCAGA 1380
Qy 1381 GACCCCAAAGGCTATGTCCAAAGCCTGCTCCGCTCCAGAGCCGGGACCTCAAGGTGATG 1440
Db |||||
Db 1381 GACCCCAAAGGCTATGTCCAAAGCCTGCTCCGCTCCAGAGCCGGGACCTCAAGGTGATG 1440
Qy 1441 ACAGATGAGCCGGCAACCTGAAGAGAGCGCGGGCTGAGTCTACACAGCCCTGG 1500
Db |||||
Db 1441 ACAGATGAGCCGGCAACCTGMAAGAGAGCGCGGGCTGAGTCTTCTACACAGCCCTGG 1500

OY	1501	TCCGAGGAGGGCGTAGTCGCATCTTACGTGCAGAATCCAGCAGGCCGAGGACGACTG	1560
Db	1501	TTCCAGGAGGCCGTGATGTCGTAATTCTACTGTCAAGATCCAGCGAGCGAGGACGACTG	1560
OY	1561	GAGCAGTCGCTGTGTTGTCCGACAACCCTGAGAGGCCAATAAACAGCAGCACGACGAAT	1620
Db	1561	GAGCAGTCGCTGTGTTGTCCGACAACCCTGAGAGGCCAATAAACAGCAGCACGACGAAT	1620
OY	1621	TTTACGCCGTGTCCGGGACCCACGACATTTTGCCCCGGGGCTCCAGCTCACTCTCTGCAAC	1680
Db	1621	TTTACGCCGTGTCCGGGACCCACGACATTTTGCCCCGGGGCTCCAGCTCACTCTCTGCAAC	1680
OY	1681	TTGGGGGTGGGGGCTGATTTAAAGTCATTCATCTGGGAATTC	1724
Db	1681	TTGGGGGTGGGGGCTGATTTAAAGTCATTCATCTGGGAATTC	1724
RESULT 2 LOCUS	BC002628	1698 bp mRNA linear PRI 12-Jul-2001	
DEFINITION	Homo sapiens, similar to SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3, clone MGC:3653 IMAGE:3608512, mRNA, complete cds.		
ACCESSION	BC002628		
VERSION	BC002628.1 GI:12803586		
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 1698)		
JOURNAL	Strausberg,R. Direct Submission Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-room1@nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nigrl.nih.gov Shaychanko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Masello,C., Mastrian,S.D., McCloskey,D.C., McQuell,J., Pearson,R., Snyder,B., Stantitop,S., Thomas,P.J., Tjongson,E.E., Touchman,J.W., Tsurgeson,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Green,E.D.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 12 Row: A Column: 13 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507086. Location/Qualifiers

```

1..1698
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="MGC:3653 IMAGE:3608512"
    /tissue_type="uterus, endometrium adenocarcinoma"
    /clone_id="NH MGC 44"
    /lab_host="DH10B-R-"
    /note="Vector: pOTB7"
    .139..1551
```

[illegible]

QY	856	GTGAAACGGCTGGGACCTGAGTGTGGGCTGCACGCTGCTCCTCATCTGACTGACACG	915
DB	817	GTGAAACGGCTGGGACCTGAGTGTGGGCTGCACGCTGCTCCTCATCTGACTGACACG	876
QY	916	CCTCCCAAGTTCAAACCTGATCCCGCTAGCCCGCTGCTGGGCTGCACACACAGAGC	975
DB	877	CCTCCCAAGTTCAAACCTGATCCCGCTAGCCCGCTGCTGGGCTGCACACACAGAGC	936
QY	976	CGCTCAGCATGTCACGCGCTGTGGCAGTATGTGAAGACCAACAGAGCTGCAGACTCC	1035
DB	937	CGCTCAGCATGTCACGCGCTGTGGCAGTATGTGAAGACCAACAGAGCTGCAGACTCC	996
QY	1036	CATGACAAAGGAATACATCAATGGGACCAAGTATTTCCAGCAGATTTTGTGATTTGCCCGG	1095
DB	997	CATGACAAAGGAATACATCAATGGGACCAAGTATTTCCAGCAGATTTTGTGATTTGCCCGG	1056
QY	1096	CTGAAGTTTCTGAGATTTCCCGCTACAGCGCTTACAGCGCTGCTATTTGCCCGCTGACCAATT	1155
DB	1057	CTGAAGTTTCTGAGATTTCCCGCTACAGCGCTTACAGCGCTGCTATTTGCCCGCTGACCAATT	1116
QY	1156	GTCATCAACCATGTCATCAGCGTGCACCGCTTACAGCGCTGCTATTTGCCCGCTGACCAATT	1215
DB	1117	GTCATCAACCATGTCATCAGCGTGCACCGCTTACAGCGCTGCTATTTGCCCGCTGACCAATT	1176
QY	1216	ATTGACGTGGAGGTGGAGGACATTAAGGGGCGAGATGAGCAGCTTCTCTATCCACG	1275
DB	1177	ATTGACGTGGAGGTGGAGGACATTAAGGGGCGAGATGAGCAGCTTCTCTATCCAGG	1236
QY	1276	GCCAAACAGCAGGAGATCAGTCTCTGGACAGTAAGATCCATGAGCAGATTTGAGTCCATA	1335
DB	1237	GCCAAACAGCAGGAGATCAGTCTCTGGACAGTAAGATCCATGAGCAGATTTGAGTCCATA	1296
QY	1336	RACGAGCTCAGATCCAGGAGCTTATGCTAAGCTTCTCAGAGACCCCAAGGCTAT	1395
DB	1297	AAACGAGCTCAGATCCAGGAGCTTATGCTAAGCTTCTCAGAGACCCCAAGGCTAT	1356
QY	1396	GTCCAAAGACCTCTCGCTCCAGAGCGGACCTCAAGGTGATGACAGATGTAGCCGGC	1455
DB	1357	GTCCAAAGACCTCTCGCTCCAGAGCGGACCTCAAGGTGATGACAGATGTAGCCGGC	1416
QY	1456	AAACCTGAAGAGAGCGCGGCTGAGTTTACCAACAGCAGCCTGTGTCAGAGAGCCGTC	1515
DB	1417	AAACCTGAAGAGAGCGCGGCTGAGTTTACCAACAGCAGCCTGTGTCAGAGAGCCGTC	1476
QY	1516	AGTCGCTACTTCTACTGCAAGATCCAGAGCGGAGCTGAGAGCTGGAGCGTGGCTGGTT	1575
DB	1477	AGTCGCTACTTCTACTGCAAGATCCAGAGCGGAGCTGAGAGCTGGAGCGTGGCTGGTT	1536
QY	1576	GTGCGCAACACCTAGGAGCCCAAAACAGCAGCAGCAGAACTTTAGCCGCTGTCCCG	1635
DB	1537	GTGCGCAACACCTAGGAGCCCAAAACAGCAGCAGCAGAACTTTAGCCGCTGTCCCG	1596
QY	1636	GGCCCCAGCATTTTCCCGGCTCCAGC-TCACCTCTCTGCACTTGGGGTGTGGGC	1694
DB	1597	GGCCCCAGCATTTTCCCGGCTCCAGC-TCACCTCTCTGCACTTGGGGTGTGGGC	1656
QY	1695	TGGATTAAGTCAATCTCTGGHAA	1720
DB	1657	TGGATTAAGTCAATCTCTGGHAA	1682
RESULT 3			
BT007694			
LOCUS			
DEFINITION			
BT007694 1413 bp mRNA linear SYN 13-MAY-2003			
Synthetic construct Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3 mRNA, partial cds.			
ACCESSION			
BT007694			
VERSION			
BT007694.1 GI:30584226			
KEYWORDS			
FLI_CDNA.			
SOURCE			
synthetic construct			
ORGANISM			
synthetic construct			
artificial sequences.			
1 (bases 1 to 1413)			
Kalinine, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S., Koundinya, M., Raphael, J., Moreira, D., Kelley, T., LaBaer, J., Lin, Y., Phelan, M. and Farmer, A.			
Cloning of human full-length CDSs in BD Creator (TM) System Donor vector			
2 (bases 1 to 1413)			
Kalinine, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S., Koundinya, M., Raphael, J., Moreira, D., Kelley, T., LaBaer, J., Lin, Y., Phelan, M. and Farmer, A.			
Direct Submission			
Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA			
This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with c-terminal tag). The CDS has been directionally cloned using BD In-Fusion (TM) cloning system between the Sali and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after Sali site and before 'ATG' to provide kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.			
Clone distribution: http://bioinfo.clontech.com/orfclones			
Location/Qualifiers			
1..1413			
/organism="synthetic construct"			
/mol_type="mRNA"			
/db_xref="taxon:32630"			
/clones="GH009113.0"			
/clone_lib="BD Creator (TM) CDS Library derived from MGC collection"			
/lab_hosts="DH5alpha T1 resistant"			
/note="Vector: pDNR-Dual"			
1..>1413			
/note="Mutations: 12:OK;1412:Stop->Leu"			
/codon_start=1			
/transl_table=11			
/product="Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3"			
/protein_id="AAP36362.1"			
/db_xref="GI:30584227"			
/translation="MTPLQHPPTVORPGMSPGARMPHOGAPMGPPSPYMGSPAVR PGLAPAGMEPARKEAPPGOSQASQOPVPTAPARSASAKRMADKILIPQRIEL VPESQAYMDLLAFERKLDQTIMKRVDIQEALEKPMQKRLRLIYINFPNPKPDAE DSDGSIAWELRVGKLLDDPSKQKRFSSFFSLVIELDKLDLYGPDNHLVWHRTPT TQETDGFQVKRPGDLSVRCTLLMLDYQPQKFLDPLRLALLGHTQSRSAIVQALWQ YVKTNRLDQSHKEYINGDKYFOQIFDCPRKFSEIPQRLTALLPDPPIVINHVIS DPSQKKTACVDIDVEVEEPLKGMSSFLSTANQOEISALDSKTHETIESINOLKIQ RDMFLSPROPKGYVQDLLRSQSRDLKVMTDVAGNPEERAEFYHQWPSQENVSRYF YCKTQRRQLEQLSVLRNTLI"			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 1403; Conservative 0; Mismatches 7; Indels 3; Gaps 3;			
QY	181	ATGACTCCAGGTCTTCAGCACCCACCCACCGTGTGTACAGCGCCCGGGATGCGTCTGA	240
DB	1	ATGACTCCAGGCTTCAGCACCCACCCACCGTGTGTACAGCGCCCGGGATGCGTCTGA	60
QY	241	GCCCGGATGCCCCACCA-GGGGGCGCCATGGGCCCCCGGGCTCCCCGTACATGGGCAGC	299
DB	61	GCCCGGATGCCCCACCAAGGGGGCGCCATGGGCCCCCGGGCTCCCCGTACATGGGCAGC	120
QY	300	CCCGCGGTGCGACCCCGGCTTGGCCCCCGCGGG-ATGGAGCGCCCGCCGCAAGCAGCAGCG	358
DB	121	CCCGCGGTGCGACCCCGGCTTGGCCCCCGGGCATGGAGCGCCCGCCGCAAGCAGCAGCG	180
QY	359	CCCGCGCCC-GGCAGAGCCAGGCACAGAGCCAGGCGCCAGCGGAGCCGCCCGCCCGCGC	417

Db	181	CCCCCCCCGGGACAGGCCAGGCCAGAGCCAGGGCCAGCGGTGCCACCGCCCCCGC	240
Oy	418	CGGAGCCGCAgTgCCAAgAGgAGgAAGTgGCTgACAAATCTTCCTCCAAAGgATTGg	477
Db	241	CGAGAGCCGCAgTgCCAAgAGgAGgAAGTgGCTgACAAATCTTCCTCCAAAGgATTGg	300
Oy	478	GAGCTgGTgTCCCCGAGTCCCAgGCTTAACTgGACCTCTTTGGCATTTGAGAGgAACTGAT	537
Db	301	GAGCTgGTgTCCCCGAGTCCCAgGCTTAACTgGACCTCTTTGGCATTTGAGAGgAACTGAT	360
Oy	538	CAAAcCATCATgCGGAAGGGGTGGACATCCAGAGgGCTCTGAAGAGGCCCATGAAGCA	597
Db	361	CAAAcCATCATgCGGAAGGGGTGGACATCCAGAGgGCTCTGAAGAGGCCCATGAAGCA	420
Oy	598	AAGCGGAAGCTgCGACTCTATATCTCCAACCTTTTAAcCTGCGAACTGATgCTGAG	657
Db	421	AAGCGGAAGCTgCGACTCTATATCTCCAACCTTTTAAcCTGCGAACTGATgCTGAG	480
Oy	658	GATTCCGAGCGGACgATTgCCTCTCGGAGCTACGGGTGGAGGGGAAGCTCTTGATgAT	717
Db	481	GATTCCGAGCGGACgATTgCCTCTCGGAGCTACGGGTGGAGGGGAAGCTCTTGATgAT	540
Oy	718	CCGAGCAAAcGGAAGCGGAAGTgTCTCTTCTTCAAGAGTTTgGTCACTGAGCTGGAC	777
Db	541	CCGAGCAAAcGGAAGCGGAAGTgTCTCTTCTTCAAGAGTTTgGTCACTGAGCTGGAC	600
Oy	778	AAAGATCTTTATgGCGCCTgACAAcCAcCTCGTTGAAGGCACTGGACACCCAGACCCAG	837
Db	601	AAAGATCTTTATgGCGCCTgACAAcCAcCTCGTTGAAGGCACTGGACACCCAGACCCAG	660
Oy	838	GAGACGGAcGGCTTCAGAGTGAACGGcCTGGGgACTGAAGTgTgCGCTGACGCTGCTC	897
Db	661	GAGACGGAcGGCTTCAGAGTGAACGGcCTGGGgACTGAAGTgTgCGCTGACGCTGCTC	720
Oy	898	CTCATgCTGgACtACAGcCTTCCCAgTTCAAACTgGATTCcCGCGTgAGCCCGGTgGTg	957
Db	721	CTCATgCTGgACtACAGcCTTCCCAgTTCAAACTgGATTCcCGCGTgAGCCCGGTgGTg	780
Oy	958	GGGCTGcACACAGAGAGCGCTCAgCCATTTgTCCAGGcCTGTGGCACTATgTGAAGCC	1017
Db	781	GGGCTGcACACAGAGAGCGCTCAgCCATTTgTCCAGGcCTGTGGCACTATgTGAAGCC	840
Oy	1018	AACAGGCTGcAGGACTCCCATgACAAgGAATACATCAATgGGGgACAAgTATTTCCAGAG	1077
Db	841	AACAGGCTGcAGGACTCCCATgACAAgGAATACATCAATgGGGgACAAgTATTTCCAGAG	900
Oy	1078	ATTTTGAATTTgTCCCCGCGCTGAAGTTTCTGAGATTCCCAcGCGCTCAcAGCCTGCTA	1137
Db	901	ATTTTGAATTTgTCCCCGCGCTGAAGTTTCTGAGATTCCCAcGCGCTCAcAGCCTGCTA	960
Oy	1138	TTGCCCCCTGAcCCCAATTgTCATCAACCATgTCACAGGgAACCTTCCAGACAGAAg	1197
Db	961	TTGCCCCCTGAcCCCAATTgTCATCAACCATgTCACAGGgAACCTTCCAGACAGAAg	1020
Oy	1198	AAGACAGCGTgCTATgACATTGAcCTGAGAGTgTGAAGgACCATTTAAAGGGcAGATgAGC	1257
Db	1021	AAGACAGCGTgCTATgACATTGAcCTGAGAGTgTGAAGgACCATTTAAAGGGcAGATgAGC	1080
Oy	1258	AGCTTCTCTATgTCCAGGGCCAAcCAGAGAGATCAgTCTCTTgGACAGTAAgATCAT	1317
Db	1081	AGCTTCTCTATgTCCAGGGCCAAcCAGAGAGATCAgTCTCTTgGACAGTAAgATCAT	1140
Oy	1318	GAGAGCATgTGAAGTTCATAAACCAcGTCAAGATCCAGAGGgACTTCACTAAgCTTCC	1377
Db	1141	GAGAGCATgTGAAGTTCATAAACCAcGTCAAGATCCAGAGGgACTTCACTAAgCTTCC	1200
Oy	1378	AGAGACCCCAAAAGGTATgTCCAAACCTGCTCCGCTCCAGAGCGGGACCTCAAGGTg	1437
Db	1201	AGAGACCCCAAAAGGTATgTCCAAACCTGCTCCGCTCCAGAGCGGGACCTCAAGGTg	1266
Oy	1438	ATGACAGATgTAgCCGGcAACCTTGAAGAGAGCGCGGGCTGAATTCTAACACCAcCC	1497

Db	1261	ATGACAGATGTAGCCGGCAACCCTGAAGAGAGCGCGGGGTAGTTCCTAACCAACAGGCC	1380
Oy	1498	TGTTCCAGAGAGCCGCTCAGTCAGTCGTAATTCTACTTGCAAGATCCAGCAGCCGAGCAGAG	1557
Db	1321	TGGTCCCAGAGGCGCGCTCAGTCGCTACTTCTACTGCAAGATCCAGATCCAGCAGCGCAGGAG	1380
Oy	1558	CTGAGAGCATGCCTGGTGTGTGGCGCAACCTTAG	1590
Db	1381	CTGAGAGCATGCCTGGTGTGTGGCGCAACCTTAG	1413
RESULT 4	BC013122		
LOCUS	BC013122	1790 bp	mRNA linear ROD 16-APR-2003
DEFINITION	Mus musculus SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3, mRNA (CDNA clone MGC:8018 IMAGE:3586276), complete cds.		
ACCESSION	BC013122		
VERSION	BC013122.1	GI:15341882	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
AUTHORS	Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heile,F., Diatchenko,L., Marusina,K., Farmer,A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ushed,T.B., Toshiyuki,S., Carninci,P., Plange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Wotley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.M., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bonfield,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schumacher,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.T., Skalska,U., Smallos,D.E., Schermer,A., Schein,J.E., Jones,S.V., and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903 (2002)	
MEDLINE	22389257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 1790)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amgc@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Louissege, H., Kowitz, C.R., Sned, A.J., Martin, R.G., Muzny, D.M., Navevati, A.N., Gibbs, R.A.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Series: IRAK project: 10 Row: 1 Column: 23

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13385379.

```

FEATURES
    source            1..1790
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="FVB/N"
                        /db_xref="taxon:10090"
                        /clone="MGC:8018 IMAGE:3586276"
                        /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                        model. 10 month old virgin mouse. Taken by biopsy."
                        /clone_lib="NCI CGAP_Mam1"
                        /lab_host="DH10B"
                        /note="Vector: pCMV-SPORT6"

    gene              1..1790
                        /gene="Smarcd3"
                        /db_xref="LocusID:66993"
                        /db_xref="MGI:1914243"
                        488..1633

    CDS                447 a 522 c 516 g 305 t
                        /product="Smarcd3 protein"
                        /protein_id="AAH13122.1"
                        /db_xref="GI:15341883"
                        /db_xref="LocusID:66993"
                        /translation="MADKILPQRIELVPEQAYMDLLAFERKLDQTIMRKRVDIQEA
                        LKRPWKQRKDLVYDNTFNPAKPAEDSDGSIASWELRVEGLKLLDDPSKQKRFSSP
                        FKSUVLELDKDLVGNHLEWHRTPTTQETDGFQVFRPGLSVRCTLMLLDYQPPQ
                        FKLDPRLARLLGLHTQSRSAIVQALWQYVKNRQLQSDHDKYINGDKYFQOIFQCPRL
                        KPSELPQRLTALLPPDPVIVNHVSVDPSDKKTACVDIVVEEPLQKGMSSFLLS
                        TANQETISALDSKSHETTESINQIKRDFMLSFSDPKGYVQDLRLSLRSQSDRLKVMTD
                        VAGNPEERRAEFVHPWSQBAVSRYFYCKIQORRQLEQSLVVRNT"

BASE COUNT          447 a 522 c 516 g 305 t
ORIGIN
Query Match          69.5%; Score 1198.6; DB 10; Length 1790;
Best Local Similarity 92.1%; Pred. No. 1.1e-217;
Matches 1297; Conservative 0; Mismatches 109; Indels 3; Gaps 3;

QY 219 GCGCCCCGGGATCGCGTCTGGAGCCCGCGATGCCACCA-GGGGCGCCATGGGCCCCC 277
DB 259 GCGCCCCGGGATGCGCGTCTGGAGCCCGAATGCCACCGGGGCGCCATGGGCCCCC 318
QY 278 GGGCTCCCGTACATGGGAGCCCGCGTGGACCCGCGCTGGCCCGCCCGGGG-ATGGA 336
DB 319 GGGCTCCCGTACATGGGAGCCCGCGTGGACCCGCGCTGGCCCGCCCGGGGATGA 378
QY 337 GCGCCCGCGCAAGCGAGAGCGCCCGCCCGCC-GGAGAGCCAGGACAGAGCGAGGCCA 395
DB 379 GCGCCCGCGCAAGCGAGAGCGCCCGCCCGCGCGGAGAGCCAGGCGAGGGCCA 438
QY 396 GCGGAGCCCGCCCGCCCGCGGAGCGCGAGTSCCAAGAGGAGGAAGATGGCTGACAA 455
DB 439 GCGCGTGGCCCGCCCGCCCGCGGAGCGCGAGTSCCAAGAGGAGGAAGATGGCTGACAA 498
QY 456 AATCTCTCTCAAAGGATTCGGGAGCTGGTCCCGAGTCCCGAGTCCCGAGGCTTACATGGACCTCT 515
DB 499 AATCTCTCTCAAAGGATTCGGGAGCTGGTACCGAGTCCCGAGTCCCGAGGCTTACATGGACCTCT 558
QY 516 GGCATTTGAGAGGAATCGATCAAAACCATCATATGCGGAAGGGGTGGACATCCAGAGGC 575
DB 559 AGCATTTGAGAGGAATCGATCAAAACCATCATATGCGGAAGGGGTGGACATCCAGAGGC 618
QY 576 TCTGAAGAGGCCATGAGCAAAAGCGGAAGCTGAGCTCTATATCTCAACACTTTAA 635
DB 619 CTTGAAGAGGCCATGAGCAAAAGCGGAAGCTGAGCTCTATATCTCAACACTTTAA 678
QY 636 CCTCTCGAAGTCTGATGCTGAGGATTCGAGCGGAGCATTTGCTCTCTGGGAGCTACGGGT 695
DB 679 CCTCTCGAAGTCTGATGCGGAGACTCTGTGTCGAGCATTTGCTCTCTGGGAGCTGCGGT 738
QY 696 GGAGGGGAAGCTCTTGGATGATCCAGCAAAACAGAAAGCGGAAGTTCTCTTCTTTCAA 755
DB 739 GGAGGGGAAGCTCTTGGATGATCCCTAGTAAGCAGAAAGGAAGTTTCTTCTTCTTCAA 798

```

RESULT 5
BC026783
LOCUS BC026783
DEFINITION Mus musculus, clone IMAGE:4481097, mRNA, partial cds.
ACCESSION BC026783
VERSION BC026783.1
KEYWORDS GI:20073202
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

BC026783 3086 bp mRNA linear ROD 13-JAN-2003
Mus musculus, clone IMAGE:4481097, mRNA, partial cds.

REFERENCE 1 (bases 1 to 3086)

AUTHORS Strausberg, R.
TITLE Direct Substitution
JOURNAL Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: <http://image.llnl.gov>
Series: IRAX Plate: 41 Row: h Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13937376.

FEATURES

source

1..3086

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:4481097"

/tissue_type="Eye, retina, mouse strain C57BL/6"

/clone_1lb="NIH_MGC_94"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

<1..1495

/codon_start=2

/product="Unknown (protein for IMAGE:4481097)"

/protein_id="AAH26783.2"

/db_xref="GI:27693623"

/translation="SGAGVAAALPGCTPGPVRMGAPGGLYRSPPWGAAYVRPG
MLGRMTPOGSMGPBGYGNPVRVPLAAGSMOSRRPAPPOIOVOOAOVONRN
HNAKKKADKILPDRIRRELVEESQAYMDLAFERKLDOTIRKSLDIDELAKRRIKO
KRLKIFISNTFNPAKSDAEDGTVASVELRVEGLDEALSKYDARKQKRSSE
FSLVIELDKDLYGPDNLVEMWRTATDETGFQVKGEDVAVCTVLMDDYOPQ
FLDPLRLARLGIHTQTRVILIOALMOYIKTHLQDPHREFFVLDKYLQOIFESQRM
KFSBIPORLHALMPEPIIINHVISVDNDOKTKACYDIEVDVDTLKTOMNSFLS
TASOEIATLNDKIHETITETINOLKTQREPMCSFAPDPGFIINDMLOSQRDLKMTMD
VGNPBEERAEFIQPMVMOEAVCRFYFSKVQRORELEALGIRNT"

CDS

BASE COUNT 680 a 911 c 798 g 697 t
ORIGIN

Query Match 40.7%; Score 701.2; DB 10; Length 3086;
Best Local Similarity 71.3%; Pred. No. 3.8e-123;
Matches 1000; Conservative 0; Mismatches 358; Indels 44; Gaps 4;

217 CAGCGCCCGGAGTGCCTCTGAGGCCCGAGTGCACCCAGGCGGCCCATGGGCCCC 276
Db CGAGACCAAGTATGTGCGCAGGTAGCGAATGACACCTCAGGGAACCTTCATGGACCT 181
Qy CGGGCTCCCGTACATGGGAGCGCCGCGCTGGGACCGGCTGGCGCCCGC-GGGATGG 335
Db C-----CTGGCTATGGGGGAAACCTTCAGTCCGACCTGCTGGCGCCAGTCAAGGATGG 236
Qy AGCGCGCCCGCAGCGAGCGAGCGCCCGCGCGAGAGCCAGGCAAGGCAAGGCGCA 395
Db ACCAGTCCCGCAGAGAGACTGCACCTC-----AACCAATCCAGCAGGT 279
Qy GCCGAGCCACCGCCCGCGCGAGCGAGCGAGTGCACAGAGAGAGAGATGGCTGACA 455
Db CCAGCAGCGAGCGGCTCAAAATCGAAATCACAATGCAAAAGAAAAGAGATGGCTGACA 339

Qy 456 AATCTCTCCCTCAAGAGATTGGAGACTGTGTCGCCAGTCCAGGCTTACATGACCTCTT 515
Db 340 AATCTCTCCCTCAAGAGATTGGAGACTGTGTCGCCAGTCCAGGCTTACATGATCTCT 399
Qy 516 GGCATTGAGAGAAATCGATCAACCATCATGCGGAAAGGGGTGAGATCCAGAGGC 575
Db 400 GCGTTTGAAGAGAAATCGATCAACCATCATGCGGAAAGGGGTGAGATCCAGAGGC 459
Qy 576 TGTGAAGAGCCCATGAAGCAAAACCGAAGTGGAGCTCTATCTTCAACATTTTAA 635
Db 460 CTGGAAGCGTCCATCAAGCAAAACCGAAGTGGAGCTCTATCTTCAACATTTTAA 519
Qy 636 CCTCGGAAGTGTGATGCTGAGATTCGACCGGACAGATTGCTCTGGAGACTACGGT 695
Db 520 TCCGCTAAGTGGAGCGCGAGATGGGAAAGGACGGTGGCTTCTGGAGACTCCGGT 579
Qy 696 GAGGAGGAACTCTCG-----ATGATCCGAGCAACAGAGG 734
Db 580 AGAAGCGCGCTCTGGAGAGCGCGGCTGTTCAAATATGACCCACCAAGCAAAAGAG 639
Qy 735 GAAGTTCTTCTTTCTTCTTCAAGATTTGTCTATGAGCTGCAAAAGATCTTTATGCCC 794
Db 640 AAGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 699
Qy 795 TGACAACCACTCGTGTGATGCGATCGGACACCGACCGGAGAGAGAGAGAGAGAGAG 854
Db 700 AGACAACCACTCGTGTGATGCGATCGGACACCGGACCGGAGAGAGAGAGAGAGAG 759
Qy 855 GGTGAAGCGGCTGGGAGCTGAGTGTGGCTGACAGCTGCTCTCATGCTGGATCA 914
Db 760 GGTGAAGCGGCTGGGAGCTGAGTGTGGCTGACAGCTGCTCTCATGCTGGATCA 819
Qy 915 GCTTCCCTCACTCAACTGATTCCTCCGCTACCGGCTGCTGGGCTGCAACAGAG 974
Db 820 GCGCCCGCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 879
Qy 975 CCGCTCAGCATTCGACGCGCTGGGAGTGTGAGAGACCAAGGCTGAGAGAGAGAGAG 1034
Db 880 AGCTCAGTATCATTCAGGACCTGGAGCTGTGAGATTTAAACACCAAGGCTGAGAGAG 939
Qy 1035 CCATGACCAAGAAATCATCAATGAGGAGCAAGTATTCAGCAGATTTTGAATGTCGCCG 1094
Db 940 TCAGAGGAGAGGTTGTTCTCTGTGACAGTACTCCAGAGATCTTGAATCTCAGG 999
Qy 1095 GCTGAAGTTTCTGATTCCTCCAGCGCTCAAGCCTGCTATTGCCCCCTGACCAAT 1154
Db 1000 GATGAAGTTCTGAGATTCCTCCAGCGCTCCAGCCTTGTATGCAACAGAGCCCAT 1059
Qy 1155 TGTCAACAACATGTCATCAGGCTGAGCCCTCAAGCAGAGAGAGAGAGAGAGAGAG 1214
Db 1060 CATCAATCAATGTCATCAGGCTGAGCCCAATCAAGAGAGAGAGAGAGAGAGAGAG 1119
Qy 1215 CATTGACGTGAGTGTGAGAGCCATTAAAGGGGAGATGAGCAGCTTCTCTTATCCAC 1274
Db 1120 CATTGACGTGAGTGTGAGAGCCATTAAAGGGGAGATGAGCAGCTTCTCTTATCCAC 1179
Qy 1275 GGCACCAAGCAGAGATCAGTCTCTGAGACGTAAATCCATGAGAGAGATGAGTCAAT 1334
Db 1180 TGCACAGCAGCAGAGATCAGCCTCTAGACAAAGATCCATGAGAGAGATGAGTCAAT 1239
Qy 1335 AAACAGCCTCAAGATCCAGAGAGCTTATGATAGCTTCTTCCAGAGAGAGAGAGAG 1394
Db 1240 CAACAGCCTCAAGATCCAGAGAGCTTATGATAGCTTCTTCCAGAGAGAGAGAGAG 1299
Qy 1395 TGTCCAGAGCTGCTCCGCTCCAGAGCGGAGCTCAAGGTGATGAGATGAGTACCGG 1454
Db 1300 CATCAATGATGCTTCACTGCTCAGTCCAGTCAAGAGAGAGAGAGAGAGAGAGAGAG 1359
Qy 1455 CAACCTGAGAGAGAGCGGCTGAGTCTTCAACAGAGAGAGAGAGAGAGAGAGAGAG 1514
Db 1360 TAAACCGAGAGAGAGCGGCTGAGTCTTCAACAGAGAGAGAGAGAGAGAGAGAGAG 1419

Db	1071	CATCATTTAATCATGTCTCATGAGTTTGACCCGAAATGATCAGAAAAAGACAGCTTGTATGA	1130
Oy	1215	CATTGACCTGGAGGTGGAGAGCCATTAAAGGGGACATGACAGCTTCTCTTATCCAC	1274
Db	1131	CATTGATGTTGAAGTGATGACACCTTGAAACCCAGATGATTTCTTTTCTGTCTGCAC	1190
Oy	1275	GGCCAAACGACGAGAGATCACTCCTCTTGACAGTAAATTCATGAGACGATTTGATGCAT	1334
Db	1191	TGCCAGCCMAAGGAGATTGCTACTCTTGACAAACAAGATCCATGACGCAATATGAAACCAT	1250
Oy	1335	AAACCCAGCTCAAGATCCAGAGGGACTTATCTCTAGCTTCTCCAGACACCCCAAGGCTA	1394
Db	1251	CAACCAAGTGAAGACTCAGCGGGATTCATCTCAGCTTTGCCAGAGACCTTCAGGGCTTT	1310
Oy	1395	TGTCCAAGACCTGCTCGGCTCCGAGCGGGACCTCAAGGTGATGACAGTGTAGCCGG	1454
Db	1311	CATCTAAAGACTGCGTTTCAGTCCAGTGCAGGAGACCTCAAGCAATGACTGATGTGTGGG	1370
Oy	1455	CAACCCCTGAAGAGAGCGCGGGGCTGAATTTACACACAGGCCCTTGATCCAGAGAGCCGT	1514
Db	1371	TAACTCCAGAGGAGGAGCGCCAGCTGAGTTTACTTCCAGGCCCTTGAGCTCAGGAGCGCT	1430
Oy	1515	CAGTCGCTACTTCTACTGCAAGATCCAGACGCGACGAGGAGGCTGGAGCAGTGGCTGT	1574
Db	1431	GTCGCCGATCTTCTACTCCAGAGGTGCACACAGACGACACAGAAATTAGACAAGCCCTGGG	1490
Oy	1575	TGTGCGCAACCTTAGAGCCCAAAAAA	1603
Db	1491	AATCCGGAATCATAGGGCTCTCCACCA	1519

```
QY 336 AGCCGCGCCGCAAGCAGCAGCGCCCGCCCGCGGAGAGCCAGGACAGAGCCAGGGCCA 395
Db 419 ATCAGTCCGCAAGAGACGTGCCCTCAG-----CAGATCCAGCAGGT 461
QY 396 GCCGAGCCACCGCCCGCGGAGCGGAGTCCCAAGAGGAGGAGAGATGGCTGACAA 455
Db 462 CCAGCAGCGCGGTCCAAAATCGAAACCAATGCAAAAGAAAAGAGATGGCTGACAA 521
QY 456 AATCTCTCCTCAAAAGGATTCGGGAGCTGGTCCCGAGTCCCAAGGCTTACATGGACCTCTT 515
Db 522 AATCTCTACTCAAGGATTCGTGAATCGTGAATCGTACCAGAAATCCCAAGGCTATATGGATCTCT 581
QY 516 GGCAATTCAGAGGAAACTGGATCAAAACCATCATCGGAAGGGGTGGACATCCAGAGGC 575
Db 582 GGCCTTTGAAAGGAACTGGACACGACTATCATGAGGAAACGGCTAGATATCCAAAGAGGC 641
QY 576 TCTGAAGGCCCATGAAGCAAAAGCGAAGTGGACTCTATATCTTCAACACTTTTAA 635
Db 642 CTTGAAGCGTCCCATCAAGCAAAAACGGAAGCTGCGAATTTTCATTTCTAAACACTTTCAA 701
QY 636 CCTCGGAAGTCTGATGCTGAGGATTCGCACGGAGCATTCGCTCTGGGAGCTACGGGT 695
Db 702 TCCGCTTAAGTCTAGATGCGGAGGATGGGAAGGAGCGGTGGCTTCCTGGGAGCTTCGGGT 761
QY 696 GGAGGGGAAGCTCTGG-----ATGATCCAGCAACAAAGAGCG 734
Db 762 AGAAGGACGGCTCTGGAGGATTCAGCCTTGTCCAAATATATGTCGCACTAAACAAAAGAG 821
QY 735 GAAGTTCTTCTTCTTCAAGAGTTGGTCAATCGAGCTGGACAAGATCTTTATGGGCC 794
Db 822 GAAGTCTCTTCTCTTTTAAAGTCTCTGGTGAATGGAACCTGGACAAAGACCTGTATGGGCC 881
QY 795 TGACAAACACCTGTTGATGTCGATCGGACACCCACGACCCAGGAGAGCGACGGCTTCCA 854
Db 882 AGACAACCATCTGTGTAGATGGCAGACGCCCACTACCCAGGAGACCGATGGCTTCA 941
QY 855 GGTGAACCGGCTGGGGACCTGAGTGTGCGTGCACGCTGTCTCTCATCTGCTGGACTACCA 914
Db 942 GGTGAAGCGCGCGGAGACGTGAATGTACGGGTGTACTGTCTCTACTGATGCTGGATTACCA 1001
QY 915 GCCTCCCAAGTTCAAACTGGATCCCGGCTAGCCCGGCTGCTGGGGCTGGACACAGAG 974
Db 1002 GCCTCCCAAGTTTAAATTTAGACCCCGGCTAGCTGACCTCTCTGGGACCTTCAATCCAGAC 1061
QY 975 CGCTCAGCATTGTGTCAGGCGCTGTGCGAGTATGTGAAGACCAACAGGCTGCAGAGCTC 1034
Db 1062 TGTCTCAGTGATCATCCAGCACTGTGCAATATATTAAGACACATAGCTCCAGGACC 1121
QY 1035 CCATGCAAGGAATACATCAATGGGGAACAAGTATTTCCAGCAGATTTTGTGATTTGTCCCG 1094
Db 1122 TCACGAGCGGAGTTTGTATCTGTGTAACAAGTACCTGCAGCAGATCTTTTGAGTCTCAACG 1181
QY 1095 GGTGAAGTTTCTGAGATTCGCCAGGCGCTACAGCCCTGTATTTGCCCTTGACCCUAT 1154
Db 1182 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTGTCTTATGSCCACCAGAACCTAT 1241
QY 1155 TGTCAATCAACATGTCATCAGGTGGAACCTTTCACACGAGAGAGACAGCGTGTATGA 1214
Db 1242 CATCATTAATCATGTGTCATGTTGACCCGATGATCAGAAAAGACAGCTTGTATGA 1301
QY 1215 CATTGACGTGGAGTGGAGGACCAATTAAGGGGAGATGAGCAGCTTCTCTCATTCAC 1274
Db 1302 CATTGATGTTGAAGTGGATGACACCTTGAAGACCCAGATGAATCTTTTCTGCTGTCAC 1361
QY 1275 GGCCAAACAGCAGGAGATCAGTCTCTTGAAGTAAAGTCAATGAGACGATGAGTCCAT 1334
Db 1362 TGCCAGCCAAACAGGAGATTTGCTACTCTAGACAAACAGATCCATGAGACAAATAGAAACCAT 1421
QY 1335 AAAACAGCTCAAGATCCAGAGGAGCTTCATGCTTAAGCTTCTCCAGACCCCAAGGGCTA 1394
Db 1422 CAACAGCTGAAGATCAGCGGAGTTCATGCTGAGCTTTGCCAGAGACCTCAGGGTTT 1481
QY 1395 TGTCCAAGACCTGTCTCCGCTCCAGAGCGGAGACCTCAAGGTGATGACAGATGTAGCCGG 1454
```

```
Db 1482 CATCAATGACTGGCTTCAGTCCCAAGTCCAGGAGCTCAAGACAATGACTGATGGTGGG 1541
QY 1455 CAACCTCTGAAGAGAGCGCGGCTGAGTTCTTACCACAGCCCTGGTCCCAAGGAGCGGT 1514
Db 1542 TNACCCAGAGAGAGCGCGGAGTGTGATTTCTATTTCCAGCCCTGGGCTCAGGAGGCTGT 1601
QY 1515 CAGTCGCTACTTCTTACTTCCAGATCCAGCAGCGAGGAGCTGGAGCAGTCCGCTGGT 1574
Db 1602 GTGCCGATACTTCTTACTTCCAAAGGTGCAGCAGACGACAAGAAATTAGAGCAAGCCCTGGG 1661
QY 1575 TGTCCGCAACACCTAGGAGCCCAAAACA 1603
Db 1662 AATCCGAATACATAGGGCCTCTCCCAACA 1690

RESULT 8
AX305314
LOCUS AX305314 3103 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 65 from Patent WO0188188.
ACCESSION AX305314
VERSION AX305314.1 GI:17644884
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 65 22-NOV-2001;
School Juridical Person Nihon University (JP)
FEATURES
Location/Qualifiers
1..3103
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
BASE COUNT 657 a 925 c 817 g 704 t
ORIGIN

Query Match 39.7%; Score 685.2; DB 6; Length 3103;
Best Local Similarity 71.1%; Pred. No. 4.1e-120;
Matches 997; Conservative 0; Mismatches 358; Indels 47; Gaps 5;

QY 217 CAGCGCCCGCGGATGCGCTCTGGAGCGCGGATGCCACACAGGGCGGCCATGGGCCCCC 276
Db 166 CCGAGACCGAGTATGCTGCCAGGTAGCCGATGACACTCAGGACCTTCCATGGGACT 225
QY 277 CGGGCTCCCCGTACATGGGAGCCCCCGCGTGCACCCCGGCTCGCCCCCGC-GGGATGG 335
Db 226 C-----CTGGCTATGGGGGAAACCTTCAGTCCGACCTGTCTGGCCCGAGTCAGGATGG 280
QY 336 AGCCCGCCCGCAAGCAGCAGCGCCCGCCCGCGCAGCGACGACGACGACGACGAGGCCA 395
Db 281 ACCAGTCCCGCAAGACCTGCACCTC-----AACAGATCCAGCAGGT 323
QY 396 GCGGAGCGCCACCGCCCGCGCGAGCGCGAGTGCAGTGCAGAGGAGGAGATGGCTGACAA 455
Db 324 CCACAGCAGCGCGTCCAAAATCGAAATCAATATGCAAGAAAAGAAAGATGGCTGACAA 383
QY 456 AATCCTCCCTCAAGGATTCGGGAGCTGGTCCCGAGTCCCGAGTCCAGGCTTACATGGACCTCT 515
Db 384 AATCCTACCTCAAGGATTCGGGAACTGGTCCCAAGATCAGAGCTACAGGCTTACATGGATCTCT 443
QY 516 GGCAATTCAGAGGAAACTGGATCAAAACCATCATCGGAGGGGTGGACATCCAGGAGGC 575
Db 444 GGCTTTTGAAGGAAACTGGACCAAGCTATTTATGAGGAAGCGGCTAGATATCCAGGAGGC 503
QY 576 TCTCAGAGGCGCATGAAGCAAAAGCGGAAGCTCGACTCTATATCTCCAACTTTTAA 635
Db 504 CTTGAAGAGTCCCATCAAGCAAAAACGGAAGCTCGGAATTTTCATTTCTAACACGTTCAA 563
QY 636 CCCTCGCAAGTCTGATGCTGAGGATTCGACGCGAGCATTCGCTCTCTGGGAGCTACGGGT 695
```

```

Db      564 TCCGCTAAGTCGAGCGGAGGATGGGAAAGGACGTCGCTTCTCGAGACTCGGGT 623
        696 GGAGGGGAGCTCCCG-----ATGATCCAGAAACAGAAAGG 734
Qy      624 AAAAAAGCGGCTCTGGAGAGCGCGCTTGTCCAAATATGACGCCACACAAACAAAAG 683
Qy      735 GAAGTCTCTCTCTCTCAAGATTTGGTTCATCGAGTGAACAAAGATCTTATAGGCC 794
Db      684 AAAAGTCTCTCTCTCTCTCAAGTCTCTGATCGAAGTGCAGAAAGACTTATAGGCC 743
Qy      795 TACAAACACCTGTTAGTGCATGACACCCAGACCCAGAGAGAGAGAGAGAGAGAG 854
Db      744 AGACAAACACTGTGTAGATGACACAGACCCGCACTACCCAGAGAGAGAGAGAGAG 803
Qy      855 GGTGAAGAGCGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 914
Db      804 GGTGAAGAGCGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 863
Qy      915 GCGTCCCGAGTCAAACTGGATCCCGCGCTAGCGCGGCTGCTGGGCTGACACAGAG 974
Db      864 GCGTCCCGAGTCAAACTGGATCCCGCGCTAGCGCGGCTGCTGGGCTGACACAGAG 923
Qy      975 CGCTCAGCAATGTCAGAGCGCGCTGTCAGAGTATGTAAAGACCAAGAGCTGAGAGCTC 1034
Db      924 AGCTCAGAGTATCATCAAGCACTGTGAGATATATAAACAACAAAGCTCAAGAGCC 983
Qy      1035 CCATGACAAAGAAATACATCAATGGGAGCAAGATTTCCAGAGATTTTGTATGTCCTCCG 1094
Db      984 TCACGAGCGAGAGTTGTTCTCTGTGACAAATACCTCCAGAGATCTTGAATCTCAGG 1043
Qy      1095 GCGTGAAGTTTGTGAGATCCCGAGCGCTCAAGCGCGCTGATGCGCGCTGAGAGCCAT 1154
Db      1044 GATGAAGTTCTAGAGATCCCTCAGCGCTCAAGCGCTGCTTATGTCACAGAGCCAT 1103
Qy      1155 TGTCAATCAACATGTCATCAGAGCGTGAACCTTCAGAGCAAGAGAGAGAGAGAGAG 1214
Db      1104 CATCATCATCATGTCATCAGTGTGAGCCCAATGACAGAAAGAGAGAGAGAGAGAG 1163
Qy      1215 CATTAAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
Db      1164 CATTAAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1223
Qy      1275 GCGCAACGAGAGAGAGATGATGTCCTGAGACAGTATGATGATGATGATGATGAT 1334
Db      1224 TGCAGAGCGAGAGAGATGAGCACTGTAGACAAAGATGATGATGATGATGATGAT 1283
Qy      1335 AAACAGCTCAAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
Db      1284 CAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1340
Qy      1395 TGTCCAAAGCTGCTCGCTCCAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
Db      1341 CATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1400
Qy      1455 CAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
Db      1401 TAAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1460
Qy      1515 CAGTGCCTACTTCTACTGCAAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
Db      1461 GTGCCGATCTTCTACTCCAAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1520
Qy      1575 TGTGCGCAACACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
Db      1521 AATCGAAACACATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1542

```

```

RESULT 9
LOCUS   MMU66620 3103 bp mRNA linear ROD 18-SEP-1996
DEFINITION Mus musculus SWI/SNF complex 60 KDa subunit (BAF60a) mRNA, complete cds.

```

```

ACCESSION U66620
VERSION   U66620.1 GI:1549248
KEYWORDS
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 3103)
AUTHORS   Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
JOURNAL    EMBO J. 15 (1996) In press
AUTHORS   Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.
TITLE      Diversity and specialization of mammalian SWI/SNF complexes
JOURNAL    Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE   96397413
PUBMED    8804307
REFERENCE 3 (bases 1 to 3103)
AUTHORS   Wang, W., Xue, Y., Zhou, S. and Crabtree, G. R.
TITLE      Direct Submission
JOURNAL    Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
FEATURES
source    1..3103
           /organism="Mus musculus"
           /mol_type="mRNA"
           /db_xref="taxon:10090"
           /note="similar to musd15k21, GenBank Accession Number M25773"
gene      1..3103
           /gene="BAF60a"
           /note="BAF60a"
           /note="ubiquitously expressed; similar to human BAF60a, BAF60b, BAF60c and yeast Swf73 and Swf73b"
           /product="SWI/SNF complex 60 KDa subunit"
           /protein_id="AAC52794.1"
           /db_xref="GI:1549249"
           /transcript="GAPAGGGLYRSPMPGAAYPRPMLPGSRMTPOGSPMPGPGY
           NPSVRPLAAGSDGSRKRAPOIOVOVOOAVONRNHNAKKKXADILIPORIELV
           PESQAYMLDLPERKLDOTIRKRLDIBALKRPFKQKRLRIFISNFPNPKSABD
           GEGIVASWELVEGRLEDAALSKYDARKQKRSFSLVLEKDLVGDNLVE
           WHRTATTOETDGFQYKRGDVNRCSTVLMVDYOPKLPDLRLRLGITHQBPVI
           IOALVQYKTHKLODPERFEVLCDYKIQOIFESQRMKFSFEPQLHALMLPBP111
           NHVIVSDBNDRKTCYDIDVEDDTLTKOMNSFLSTASQOEIATLNNKHETETI
           NOKTORREFMLSLPESGFINDMLQSCGRGLTKMTDVGNGNPEERAEFYQPMAQEA
           VCRIFYSVQORROBLEOALGIRNT"
BASE COUNT 657 a 925 c 817 g 704 t
ORIGIN
Query Match 39.7% Score 685.2; DB 10; Length 3103;
Best Local Similarity 71.1%; Pred. No. 4,1e-120;
Matches 997; Conservative 0; Mismatches 358; Indels 47; Gaps 5;

```

```

Qy      217 CAGCGCGCGGAGTCCGCTTGTGAGCCCGAGATGCCCAACAGAGGCGCCCATGAGGCCCC 276
Db      166 CCGAGACCAAGATGCTCGCAGGTAGCCGAATGACACCTCAGAGGACCTTCCATGGGACCT 225
Qy      277 CGGGCTCCCGTACATGGGACAGCCCGCGCTGACACCCGCGCTGGCCCGC-GGGATGG 335
Db      226 C-----CTGGCTATGTGGGGGAACCTTCAATGTCGACCTGTGCGCCAGTGTAGGAGTGG 280
Qy      336 AGCCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 395
Db      281 ACCAGTCCCGCAAGAGACCTGCACCTC-----AACAGATCCAGCAGGT 323
Qy      396 GCGGAGGCCAACCGCCCGCGCGAGAGCGGAGTGCACAGAGAGAGAGAGAGAGAGAG 455
Db      324 CCAGCAGCAGGCGGTCCAAATCGAAATCACAAATCGAAAGAGAGAGAGAGAGAGAG 383

```

```
Qy 456 AATCTCTCCCTCAAAGGATTTCCGGAGCTGGTCCCGAGTCCAGGCTTACATGGACCTCTT 515
    |||||
Db 384 AATCTCACTCAAGGATTCGGGAATCGTCCAGAAATCAAGGCTTACATGGATCTCT 443

Qy 516 GGCAATTCAGAGGAACTGGATCAAAACATCATGCGGAAGGGGTGGACATCCAGGAGGC 575
    |||||
Db 444 GGCATTTGAAAGGAACTGGACAGACTATTATAGGAAGGGGTAGATATCCAGGAGGC 503

Qy 576 TCTGAAGAGCCCATGAAAGCAAAAGCGGAAGCTGGACTCTATATCTCCACACTTTTAA 635
    |||||
Db 504 CTTGAAACGTCCTCAAGCAAAAGCAAGGATGCGAAATTTTCATTTCTTAACACGTTCAA 563

Qy 636 CCTCCGAAGTCTGATGCTGAGGATTCGACGGCAGCATGCTCCTCGGAGCTACGGGT 695
    |||||
Db 564 TCCGCTTAAGTCGACGCGGAGGATGGGAAGGAGCGGTGCTTCTCGGAGCTCCGGGT 623

Qy 696 GGAGGGGAAGCTCTCTGG-----ATGATCCAGCAAAACAGAAAGCG 734
    |||||
Db 624 AGAAGCGCGCTCCTGGAGGACGCGCCTTGTCCAAATATAGCGCCACCAAGCAAAAGAG 683

Qy 735 GAAGTTCTCTCTTCTTCAAGAGTTTGGTATCGAGCTGGACGAAGATCTTTATGGCCC 794
    |||||
Db 684 AAAGTTCTCTCTTCTTAAAGTCTCTTGGTATCGAACTGGACAAAGACCTCTATGGCCC 743

Qy 795 TGACAAACACCTCGTTGAGTGGCATCGGACACCCACGACCCAGGAGACGGACGGTTCCA 854
    |||||
Db 744 AGACAAACCATCTGTTAGTAATGGCACAGACCGCCACTACCCAGAGACCGATGGCTTCA 803

Qy 855 GGTGAACCGGCTGGGGACCTGAGTGTGCGTGCACGCTGCTCTCATCTGCTGGACTACA 914
    |||||
Db 804 GGTGAAGCGGCGCAGGAGATGTGAATGTACGGGTGTACTGCTCTGCTGATGCTGGACTACA 863

Qy 915 GCCTCCCAAGTTCAAACTGGATCCCGCGCTAGCCGCGTGTGGGGCTGCACACACAGAG 974
    |||||
Db 864 GCGCCCGGAGTTTAAATTAAGACCTCGCTGCGCTCTTGGGCACTCCATACCCAGAC 923

Qy 975 CGGCTCAGCATGTGTCAGGCGCTGTGGCAGTATGTGAAGACCAACAGGCTGCAGGATC 1034
    |||||
Db 924 AGCTCAGTATCATCTCAAGCACTGTGGCAGTATATTAAACACACAAAGCTCCAGGACCC 983

Qy 1035 CATGACAAAGGAATACATCAATGGGGACAAATTTTCCAGCAGATTTTGTGTCCTCCG 1094
    |||||
Db 984 TCACAGCGAGAGTTTGTCTCTGTGCAAGTACCTCCAGCAGATCTTTGAATCTCAGCG 1043

Qy 1095 GCTGAAGTTTCTGAGATTCACCGGCTCACAGCCTGCTATTGCCCCCTGACCCCAAT 1154
    |||||
Db 1044 GATGAAGTTCTCAGAGATCCCTCAGCGGCTCCACGCTTGTCTTATGCCCCAGAGCCCAT 1103

Qy 1155 TGTCAATCAACATGTCAATCAGCGTGGACCTTTCAGACCAAGAAAGACAGCGTGTATGA 1214
    |||||
Db 1104 CATCATCAATCATGTCAATCAGTGTGGACCCAAATGACCAAGAAAGACCGGTGTATGA 1163

Qy 1215 CATTGACGTGGAGGTGGAGGCCATTAAGGGGCGAGATGAGCAGCTTCTCTCTATCCAC 1274
    |||||
Db 1164 CATTGACGTGGAGGTGGATGACACTCTGAAGACCCAGATGAATCTTCTCTGTTGTCAC 1223

Qy 1275 GGCCAAACAGCAGGAGATCAGTCTCTGGAAGTAAAGTCCATGAGACGATGAGTCCAT 1334
    |||||
Db 1224 TGCCAGCCAGCAGGAGATCGCCACTCTAGACAAAGATCCATGATGAGCATAGAGACCAT 1283

Qy 1335 AAACAGCTCAAGATCCAGAGGAGCTTCATGCTAAGCTTCTCCAGACACCCCAAGGGCTA 1394
    |||||
Db 1284 CAAACAGCTGAAGACCCAGAGAGTTTCATGTTGAGCTTGCCC---GAGCCCTCAGGTTT 1340

Qy 1395 TGTCCAAGACCTGCTCCGCTCCAGAGCGGGACCTCAAGGTGATGACAGATGTAGCCGG 1454
    |||||
Db 1341 CATCAATGATTGGCTTTCAGTCCAGTGCAGGGGCTCAAGACGATGATCTGTTGGTGG 1400

Qy 1455 CAACTCAAGAGGAGCGCGGCTGAGTTCTTACCACAGCCCTGGTCCCAGGAGCGCT 1514
    |||||
Db 1401 TAAACCGGAAGAGGAGCGTCTGCTGAGTTCTTACTTCCAGCGCCCTGGGCTCAGGAGCGT 1460
```

```
Qy 1515 CAGTCGCTACTTCTTACTGGAATCCAGCAGCGCAGGAGCTGGAGCAGTGCCTGGT 1574
    |||||
Db 1461 GTGCCGATACTTCTTACTCCAAAGGTGCAGAGCGCGCAAGAGTTAGAGCAAGCCCTGG 1520

Qy 1575 TGTGCGCAACACCTAGGAGCCC 1596
    |||||
Db 1521 AATCCGAACACATAGGCGCTC 1542

RESULT 10
BC049347
LOCUS          BC049347          2747 bp      mRNA      linear      VRT 31-MAR-2003
DEFINITION    Danio rerio, clone MGC:55436 IMAGE:2639578, mRNA, complete cds.
ACCESSION     BC049347
VERSION       BC049347.1  GI:29387083
KEYWORDS      MGC.
SOURCE        Danio rerio (zebrafish)
ORGANISM      Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE     1. (bases 1 to 2747)
              Strausberg, R.
              Direct Submission
              Submitted (28-MAR-2003) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
              Contact: MGC help desk
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
              cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Institute for Systems Biology
              contact: amadan@systemsbiology.org
              Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
              Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAP Plate: 101 Row: 1 Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1..2747
/organism="Danio rerio"
/mol_type="mRNA"
/strain="AB"
/db_xref="taxon:7955"
/clone="MGC:55436 IMAGE:2639578"
/tissue_type="Whole body, adult, (one male and one female,
including unfertilized eggs)"
/clone_lib="Sugano kawakami zebrafish DR"
/lab_host="DH10B"
/notes="Vector: pME18S-FL3"
28..1560
/codon_start=1
/product="Unknown (protein for MGC:55436)"
/protein_id="AAH49347.1"
/db_xref="GI:29387084"
/translation="MAARSGFQPSPTGGGAGPMGPPPVAGAGPMGPGTSGRMGPG
GPQNHMTRSPMPYPRGMPMPASRTPGPSMPPGYGASVPSPGMPVMPSPRKRPA
PNOIQQQQNNRNOHAKKMWADKILPQRIRELPEQAYMDLLAFERKLDOTNRKR
LDIQAELKRIKQKRLRIFISNTFNPAKPADEGEGTASWELRVEGRLEDITAVSK
YEATQKRRKSSFFKSLVIDLDLYGPDNHLVEMHRTATTQETDGFQVKRPGDGVK
CTVLLMDYDPPQYKLPRLARLGIHTQTPVILQALWQVYKHLQDPHERBFNC
DKYLQIIFETQRMKFSIPLORLHALLMPEPIINHVLSDPNQKKTACACDVIDEV
DTLQIWNSSLSTASQEIAGLQNHETIETINQLKTRFMLSDFQGFINDW
LQSCDRDLTKMTDVGVPBERRAEFYQPWQAQBAVCVFYSYKQVRQEQLEALGIR
NT"
```



```
Matches 996; Conservative 0; Mismatches 357; Indels 49; Gaps 7;
Qy 217 CAGGCCCGCGGATGCGCTCTGGAGCCCGGATGCCACACAGGGGCGCCATGGGCCCCC 276
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 166 CCGAGACACAGGATGCTGCCAGGTAGCGAATGACACCTCAGGACCTTCTCCATGGGACT 225
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 277 CGGGCTCCCGTACATGGCGAGCCCGCGCTGCGACCCCGGCTGGCCCCCGC-GGGATGG 335
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 226 C-----CTGGCTATGGGGGGAACCTTCAGTCCGACCTGGTCTGGCCCGCAGGATGG 280
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 336 AGCCCGCCGCAAGGACAGGGCCCGCGCGGAGCCAGGACACAGAGCCAGGGCCA 395
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 281 ACCAGTCCCGAAGAGACCTGCACCTC-----AACAGTCCAGCAGGT 323
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 396 CGCGAGAGCCACCGCCCGCGGAGCGCGAGTCCCAAGAGGAGAGATGGGTGACAA 455
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 324 CCAGCAGCAGGGGTCGAAATCGAAATCAATGCAAGAAAGAAAGATGGCTGACAA 383
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 456 AATCCTCCCTCAAGGATTCGGGAGCTGGTCCCGAGTCCAGGCTTACATGGACCTTT 515
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 384 AATCCTACCTCAAGGATTCGGGAATCGTCCAGAAATCAAGGCCCTACATGGATCTCT 443
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 516 GGCATTTGAGAGGAACTGGATCAAAACCATCATGCGGAAGGGGTGGACATCCAGGAGC 575
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 444 GGCCTTTGAAAGGAACTGGACAGACTATTATGAGGAAGCGGTAGATATCCAGGAGGC 503
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 576 TCTGAAGAGGCCATGAAGCAAAAGCGGAAGCTGCGACTATATCTCAACACTTTTAA 635
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 504 CTTGAAACGTCCCATCAAGCAAAACGGAAGCTGCGAATTTTCATTTCTAACAGTTCAA 563
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 636 CCCTCGAAGTCTGATGTGAGGATTCGCGAGCGAGCATTCCTCTCGGAGCTACGGGT 695
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 564 TCCGCTTAAGTCGAGCGCGGAGGTGGGAAGGACGGTGGCTTCTCGGAGCTCCGGGT 623
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 696 GGAGGGGAAGCTCTGG-----ATGATCCAGCAACAGAAAGCG 734
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 624 AGAAGCGCGGCTCTCGGAGGACGCGCCCTTGTCAAATATACGCGCACCAAGCAAGAG 683
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 735 GAAGTCTCTCTTTCTTCAAGAGTTTGTGTCATCGAGCTGGCAAGAGATCTTTATGGCCC 794
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 684 AAGTTCTCTCTTTTAAAGTCTTTGATGATGATGATGATGATGATGATGATGATGATG 743
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 795 TGACAACCACTCTGTGATGTCATCGGACACCCAGCAGGAGAGCGGCTTCCA 854
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 744 AGACAACCATCTGTGTAGATGSCACAGACCGCCACTACCCAGGAGACCGATGGCTTCCA 803
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 855 GGTGAACCGGCTGGGGACCTGAGTGTGGCTGACGCTGCTCTCATCTGCTGGATACCA 914
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 804 GGTGAAGCGGCGCAGGAGATGTGAATGTACGGGTGTACTGTCTGTCTGTGCTGGACTACCA 863
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 915 GCCTCCCGAGTTCAAACTGGATCCCGCTAGCCCGGCTGTGGGGCTGCACACACAGAG 974
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 864 GCCCGCCAGTTAAATTAGACCTTCGGCTCGGCTCTTGGGCATCCATACCCAGAC 923
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 975 CCGCTCAGCCATGTCTCCAGGCGCTGTGGCAGTATGTGAAGCAACAGGCTGCAGGACTC 1034
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 924 ACGTCCAGTATCATCCAGCACTGTGCGAGTATATTAACACACAAAGCTCCAGGACCC 983
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1035 CCATGAAGGAATCATCAATGGGGCAAGTATTTCCAGCAGATTTTGTATGTCCCG 1094
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 984 TCACGAGCGAGAGTTGTCTCTGTGCAAGATGACCTCCAGCAGAGATCTTTGAATCTCAGCG 1043
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1095 GCTGAAGTTTCTGAGATTCGCCAGCGCTCACAGCCCTGCTATTGCCCCCTGACCCCAAT 1154
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1044 GATGAAGTTCTCAGAGATCCCTCAGCGGCTCCAGCCCTTGTATTGCCACAGAGCCCAT 1103
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1155 TGTCTAACCATGTCTCAGCGTGGACCTTTTCAGACCAGAGAAAGACAGCGTGTCTATGA 1214
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1104 CATCATCAATCATGTCTCAGTGTGGACCCCAATGACCAGAAAAAGACCGGTGTCTATGA 1163
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1215 CATTGACGTGGAGTGGAGGACCAATTAAAGGGGAGATGAGCAGCTTCTCTCTATCCAC 1274
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1164 CATTGACGTGGAGTGGATGACACTCTGAAGACCCAGATGAATCTTTCTCTGTGTTCCAC 1223
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Qy 1275 GGCCAAACAGCAGAGATCAGTCTCTGGACAGTAAGATCCATGAGACGATTGAGTCCAT 1334
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1224 TGCCAGCCAGCAGAGATGCCACTCTTAGACAAACAAGATCCATGAGACGATAGAGCCAT 1283
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1335 AAACACACTCAGATCCAGAGGACCTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGCTA 1394
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1284 CAACCACTGAAGACCCAGCAGAGATTCATGTTAGCTTGGCC---GAGCCCTCAGGTTT 1340
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1395 TGTCAAGACCTGCTCGCTCCCAAGAGCCGGAGCTCAAGGTGATGACAGATGTAGCCGG 1454
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1341 CATCAATGATTGGCTTCAGTCCCAAGTCAGGG-CCTCAAGACGATGACTGATGTGGGG 1399
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1455 CAACCTCTGAAGAGAGCGCGGCTGAGTTCTACCAACAGCCCTGCTCCAGGAGGCGGT 1514
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1400 TAA-CCGGAAGAGAGCGTCTGCTGAGTTCTACTTCCAGCCCTGGCTCAGGAGGCTGT 1458
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1515 CAGTCGCTACTTCTACTGCAAGATCCAGCAGCGCAGGAGGAGCTGAGCAGTGCCTGGT 1574
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1459 GTGCCGATACTTCTACTTCAAGGTGCAGAGCGGCAAGAGTTAGAGCAAGCCCTGGG 1518
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1575 TGTGCGCAACACCTAGGAGGCC 1596
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1519 AATCCGAACACATAGGCGCTC 1540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 12
AB003504
LOCUS 2297 bp mRNA linear ROD 05-FEB-1999
DEFINITION Rattus norvegicus mRNA for BAF60b, partial cds.
ACCESSION AB003504
VERSION AB003504.1 GI:2723483
KEYWORDS BAF60b.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (sites)
AUTHORS Nomoto, K., Nakazato, S., Kazahari, K. and Ono, M.
TITLE Gene structure of rat BAF60b, a component of mammalian SW1/SNF
complexes, and its physical linkage to the growth hormone gene and
transcription factor SUG/proteasome p45 gene
JOURNAL Gene 202 (1-2), 157-165 (1997)
MEDLINE 98087432
PUBMED 9427560
REFERENCE 2 (bases 1 to 2297)
AUTHORS Ono, M.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-1997) Masao Ono, Rikkyo University, College of
Science, Department of Chemistry, Life Science Course;
Nishi-Ikebukuro, 3-34-1, Toshima-ku Tokyo 171-8501, Japan
(E-mail: monoo@rikkyo.ac.jp, Tel:+81-3-3985-2387,
Fax:+81-3-5992-3434)
COMMENT Rat BAF60b is closely related to human BAF60b (Genes & Dev. 10,
2117-2130, 1996).
FEATURES
Location/Qualifiers
1..2297
/mol_type="mRNA"
/strain="Wistar"
/db_xref="taxon:10116"
/clone="HL2"
/sex="male"
/tissue_type="liver"
<1..1411
/codon_start=2
/product="BAF60b"
/protein_id="BAA24105.1"
/db_xref="GI:2723484"
/translation="GPAAPAAQYORPGMPSGRMPAGLQVPPAGSPFGTAAPLRPG
MPFTMDPFRRLLLVPAQPFMPAPRGLKRRKWADKVLPORIRELVPESQAYMDLLA
FERKLDQTIARKRMEIOEAIKKPLTKRKLRIYISNTFSPSKADGDSGTAGTGGTTP
CDS
1..2297
```


ACCESSION U66618
 VERSION GI:1549244
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 2041)
 Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Muchardt, C., Kaipana, G. V., Goff, S. P., Yaniv, M., Workman, J. L., and Crabtree, G. R.
 Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
 EMBO J. 15 (1996) In press
 REFERENCE 2 (bases 1 to 2041)
 Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R., and Crabtree, G. R.
 Diversity and specialization of mammalian SWI/SNF complexes
 Genes Dev. 10 (17), 2117-2130 (1996)
 JOURNAL MEDLINE 96397413
 PUBMED 8804307
 REFERENCE 3 (bases 1 to 2041)
 Wang, W., Xue, Y., Zhou, S., and Crabtree, G. R.
 Direct Submission
 Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
 JOURNAL TITLE Location/Qualifiers
 FEATURES
 source
 1..2041
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_type="Jurkat T cells"
 1..2041
 /gene="BAF60b"
 /gene="BAF60b"
 429..1856
 /note="highly expressed in pancreas; present in a distinct complex from BAF60a; similar to human BAF60a, BAF60b, BAF60c and yeast SWP73 and SWP73b"
 /codon_start=1
 /product="SWI/SNF complex 60 kDa subunit"
 /protein_id="AAC50696.1"
 /db_xref="GI:1549245"
 /translation="MLPGPALRGPPAAYORPQMSPNRMPARLARGPCWLPIMCS
 SASATMHPHNDGSIPTKPPALRGPAPYORPQMSPNRMPARLARGPCWLPIMCS
 YNDLAFERKLDQTIARKMEIOBAIKKLTORRKRITISNTPFSKABGSAGTGA
 TPGGTPAGGKVAWSLELVGKLDLDPSSKQKRFSSFKSLVLEDELTPRLRLGYHTPRAIM
 YMMPTQEDTGFQVKKRPGDLNVKCTLLMLDHPQPKYKLDPRILRLGYHTPRAIM
 QALMLYIKHNQDQHERREYINCNRFYRQIFSGRLRFSIEMKLAGLGHDPDIYIN
 HISVDPNDOKKTACYDIDVEYDPLKQAMSNPLASTTNOGLASLDVRIHETISIN
 OKTORPMLSTSDPQDFIQEWLSQRDLKITIDVTGNPEERRAATYHQPMQAOEA
 VGRHIFAKVQQRQELQVLRILT"
 BASE COUNT 450 a 605 c 610 g 376 t
 ORIGIN
 Query Match 36 %; Score 634.4; DB 9; Length 2041;
 Best Local Similarity 70.6%; Pred. No. 1.9e-110;
 Matches 876; Conservative 0; Mismatches 331; Indels 33; Gaps 1;
 Oy 391 GGCCAGCCGAGCCACCGCCCCCGCGAGCGCGAGCGGAGGAGGAGAGATGAGCT 450
 Db 624 GCCCCAGCCGAGCTCCCATGCTCGCCAGCGCGCGGGGTTAAAGAGGAGATGGCA 683
 Oy 451 GACAAATCTCTCTCAAAGGATCGGGAGCTGGTCCCGAGTCCAGGCTTACATGAC 510
 Db 684 GATTAAGTTCTACTCAGGGAATCCGGGAGCTTGTTCAGAGCTCAGCGTACATGAGT 743
 Oy 511 CTCTTGAGATTGAGAGAACTGATCAACCATCATGCGGAGGAGGGGGGATCCAG 570
 Db 744 CTCTTGAGCTTTGAGGAGAGCTGAGCAGACCATTCCTCGCAGGAGATGAGATCCAG 803
 Oy 571 GAGGCTCTGAGAGGCGCATGAGCAAAAGCGGAGCTGCACTTATATCTTCAACACT 630
 Db 804 GAGGCGCATCAAAAGCTCTGACACAAAAGCGAAAGCTTCGATCTACATTTCCAAATACG 863

Oy 631 TTTAACCTGCGAAGTCTGATGAGAT----- 660
 Db 864 TTCAATGCCAGCAAGCGGAAGCGATGATGACAGAACTGACAGGACCCCTGGGGAAACC 923
 Oy 661 ---TCGACGCGAGATTGCTCTTGAGAGTACGCGTGGAGGGAAGCTCTGATGAT 717
 Db 924 CCAGCAGGGGACAGAGGTGCTTCTGGGAACTCCGATGGAGGAAACATGCTGGATGAT 983
 Oy 718 CCAGCAACAGAGCGGAAGTTCTTCTTCTTCAAGATTTGGTCACTGAGCTGGAC 777
 Db 984 CCTAGCAACAGAGGAAAGTTTCTTCAATCTTTAAAGCTCGTCAATGAGCTGGAC 1043
 Oy 778 AAGATCTTTAGGCGCTGACAAACACCTCGTTGATGGCATCGGACCCGACGACCGAG 837
 Db 1044 AAGAGCTGTAAGGCTTGAACGCTGACGCTGATGAGTGTATTTGATGCTCACACCCAG 1103
 Oy 838 GAGACGAGCGGCTTCAGGTGAACCGGCTGGGGAAGCTGAGTGTGCGCTGCAAGCTGCTC 897
 Db 1104 GAGACGATGGCTTCCAGTAAAGGCGCTGGAGACCTCAAGTCAAGTGCACCCCTCTG 1163
 Oy 898 CTCATGCTGGAATCAAGCTCTCCAGTTCAAATGGAATCCCGCTAGCCGGCTGCTG 957
 Db 1164 CTCATGCTGGAATCAAGCTCTCCAGTTCAAATGGAATCCCGCTAGCCGGCTGCTG 1223
 Oy 958 GGGCTGCAACAAGAGCGGCTCAGGCAATTCAGGCGCTGAGCATGATGAGACG 1017
 Db 1224 GAGTGCAACAAGAGCGGCTCAGGCAATTCAGGCGCTGAGCATGATGAGACG 1283
 Oy 1018 AACAGCTGCAAGAGCTCCCATGACAGGATATCATCAATGAGGAGCAAGATTTCCAGCAG 1077
 Db 1284 AACAGCTGCAAGAGTGGGACAGAGGGAGTATCATCAATGAGGAGCAAGATTTCCAGCAG 1343
 Oy 1078 ATTTTGAATGTCCTCCGCTGAGATTTTCTGAGATTTCCAGGCTCAGAGCTGCTGA 1137
 Db 1344 ATCTTCAGTTGTCGCGGACTCCGTTTCTCGAGATTTCCATGAGCTGAGGCTGCTG 1403
 Oy 1138 TTGCCCCGCAAGCAATTCATCAACATGATCAGGCTGAGACCTTCAGACAGAG 1197
 Db 1404 CAGCATTCAGACCAATTCATCAACATGATCAGGCTGAGACCTTCAGACAGAG 1463
 Oy 1198 AAGACGCTGCTATGATCAATTCAGGAGTGAAGGAGCCATTAAGGGGAGATGAGC 1257
 Db 1464 AAGACGCTGCTATGATCAATTCAGGAGTGAAGGAGCCATTAAGGGGAGATGAGC 1523
 Oy 1258 AGCTTCTCTATCAACGCGCAACAGAGAGATCACTCTGAGACGTAAGATCAT 1317
 Db 1524 AATTTCTGAGCTTACCAACATCAGAGGAGATCGCTCCTTGATGTCAAGATCAT 1583
 Oy 1318 GAGAGATGATCCATTAACAGAGTCAAGATTCAGAGGAGCTTATGATGATGCTTC 1377
 Db 1584 GAGACCATGATTCATCAACAGAGTCAAGAGTCAAGATTCATGATGATGCTTC 1643
 Oy 1378 AAGACCCCAAGGCTATGCTCAAGACCTGCTCCGCTCCAGAGCGGAGCTCAAGGTG 1437
 Db 1644 ACCGACCCCAAGGCTATGCTCAAGAGCTGCTCCGCTCCAGAGCGGAGCTCAAGGTG 1703
 Oy 1704 ATCACTGATGATCAAGAAATCTGAGAGGAGAGACGAGCTGCTTTCACACAGCC 1763
 Db 1498 TGGTCCCAAGGAGCGCTGATGCTTCTTCTTCAAGATTTGGTCACTGAGCTGGAC 1557
 Oy 1764 TGGGCCCCAGGAGAGTGAAGGAGCAATCTTTCAGAGGTGACAGGAGGAGGAA 1823
 Db 1558 CTGAGCAGTCTGCTGTTGCGCAACACTTGAAGGCCA 1597
 Oy 1824 CTGGAACAGTCTGGAATTTGCTGACCTTAATCTGCTCA 1863

RESULT 15
 BC045009
 LOCUS BC045009 2557 bp mRNA linear VRT 10-FEB-2003

Db	1310	GGTAATCCTGAAGAGACGTCGAGCTGAATTTTACTTCCAAACATGGGCTCATGAGCT	1369
Qy	1513	GTCAGTCGCTACTTCTACTGCAAGATCCAGACGCGCAGGAGCTGGAGCAGTCGCTG	1572
Db	1370	GTAATGCCGGTATTTCTACTCCAAAGTGCAAGCAAAAGAGCTCAGGAGCTGSAACAGGCTTC	1429
Qy	1573	GTTGTGGCGCACACCTTA	1589
Db	1430	GGGATCCGAAATACATA	1446

Search completed: October 11, 2003, 10:35:12
Job time : 4260 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 08:22:11 ; Search time 360 Seconds
(without alignments)
12927.309 Million cell updates/sec

Title: U66619
Perfect score: 1724
Sequence: 1 GAATTCGGCGCGAGCGCC.....GTCATTCACTCGAATTC 1724

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_19jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	77.2	1452	25	AAD53119 Human BRG1-associ
2	1322.6	76.7	3089	22	AAH98321 Human EST-derived
3	1322.6	76.7	3272	22	AAH98319 Human EST-derived
4	983	57.0	1970	24	ABO54477 Human ovarian anti
5	821.6	47.7	6809	25	ABZ35936 Human secretory po
6	685.2	39.7	3103	24	AB199242 Mouse ischaemic co
7	642	37.2	2511	24	AB573872 Human cDNA encodin
8	572.6	33.2	2010	23	ABU02583 Drosophila melanog

9	511.4	29.7	2581	25	ABX34564 Human mdtl cDNA SE
10	504.2	29.2	4138	23	ABL02582 Drosophila melanog
11	466.2	27.0	755	25	AAD53120 Human BRG1-associ
12	410.4	23.8	450	24	ABT07117 Human ovarian canc
13	410.4	22.8	450	25	ABX72995 Human ovarian carc
14	206.2	12.0	285	25	ABX47225 Bovine EST associa
15	205	11.9	452	21	AA280561 Human colon cancer
16	179.2	10.4	690	22	AA280561 Human colon cancer
17	170.4	9.9	8788	22	AA135930 Human muscleoskele
18	170.4	9.9	8788	22	AA162919 Human genomic DNA
19	170.4	9.9	8788	25	ABX58918 cDNA encoding nove
20	167.2	9.7	187	16	AAT25158 Human gene signatu
21	137.8	8.0	310	22	AAF7989 Nucleotide sequenc
22	132	7.7	520	22	ABA29582 Probe #8048 for ge
23	132	7.7	520	22	AAK10561 Human brain expres
24	132	7.7	520	22	AA117316 Probe #7249 for ge
25	132	7.7	520	22	AA142206 Probe #10892 used
26	132	7.7	520	24	ABSI0458 Human genome-deriv
27	129.2	7.5	219	22	ABA39473 Probe #17939 for g
28	129.2	7.5	219	22	AAK23257 Human brain expres
29	129.2	7.5	219	22	AA126536 Probe #16469 for g
30	129.2	7.5	219	22	AA155290 Probe #23976 used
31	129.2	7.5	219	22	ABX22909 Human genome-deriv
32	110	6.4	95223	21	AAF22282 BAC containing rep
33	95.6	5.5	3824	23	ABAI03782 Drosophila melanog
34	91.6	5.3	15500	22	ABAI5840 Human nervous syst
35	91.6	5.3	15500	22	ABZ73784 Human cardiovascular
36	91.6	5.3	15500	25	ABZ73784 Secreted protein g
37	91.6	5.3	15500	25	ABT16889 Human secreted pro
38	91.6	5.3	15500	25	ABZ67378 Human secreted pro
39	84.6	4.9	2017	21	AAF21750 Human breast and o
40	67.8	3.9	215	24	ABL86508 Human ovarian can
41	66.2	3.8	740	24	ABO54483 Arabidopsis thalia
42	64	3.7	114955	20	AAK53491 Human adenosine A1
43	60	3.5	60	24	ABN33431 Human spliced tran
44	59.4	3.4	114955	20	AAK53491 Human adenosine A1
45	57.2	3.3	335	25	ABT21717 Breast cancer mark

ALIGNMENTS

RESULT 1
ID AAD53119 standard; DNA; 1452 BP.

AC AAD53119;

DT 28-MAY-2003 (first entry)

DE Human BRG1-associated factor (BAF) 60c2 DNA.

Human; nuclear receptor cofactor; proliferator-activated receptor; RXR;
PPAR-gamma; retinoid X receptor; cell proliferative disorder; infection;
metabolic disorder; cardiovascular disorder; inflammatory disease; acne;
neurodegenerative disease; Parkinson's disease; psoriasis; cancer; HIV;
renal disease; atherosclerosis; Alzheimer's disease; diabetes; vaccine;
osteoporosis; human immunodeficiency virus; BRG1-associated factor; BAF;
gene therapy; gene; ds.

OS Homo sapiens.

Key Location/Qualifiers
CDS 1..1452
FT /tag= a
FT /product= "BAF60c2 protein"

PN WO200294877-A2.

PD 28-NOV-2002.

PF 23-MAY-2002; 2002WO-IB02939.

XX

PR 23-MAY-2001; 2001US-292526P.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Auwerx J, Gelman L, Debril M, Rocchi S, Picard F;
XX WPI; 2003-120789/11.
DR P-PSDB; AAE34718.
XX
PT New human PPAR or RXR cofactor polypeptide, useful for identifying
PT modulators of nuclear receptor activity, and for treating or preventing
PT pathologic conditions associated with cell types that express PPAR
PT receptors, e.g. cancer
XX
PS Example 1; Column 55; 31pp; English.
XX
CC The invention relates to human nuclear receptor cofactor polypeptides.
CC The invention particularly relates to proliferator-activated receptor
CC (PPAR)-gamma and retinoid X receptor (RXR) cofactor polypeptides.
CC Polypeptides of the invention are useful for screening compounds that
CC modulate the interaction of the nuclear receptor with the nuclear
CC receptor cofactor. Modulators of the invention are useful in preparing
CC a pharmaceutical composition for treating and/or preventing diseases
CC or pathologic conditions associated with cell types that express PPAR
CC receptors. The pathologic conditions treated include metabolic or cell
CC proliferative disorders such as diabetes, cardiovascular disorders
CC (e.g. atherosclerosis), renal diseases, neurodegenerative diseases
CC (e.g. Parkinson's disease, Alzheimer's disease), inflammatory diseases
CC (e.g. psoriasis, acne), wounds, osteoporosis, infections (e.g. HIV)
CC or cancer. The invention is useful in gene therapy and as vaccines.
CC The present sequence is human BRG1-associated factor (BAF) 60c2 DNA.
XX This sequence is used in the exemplification of the invention.
SQ Sequence 1452 BP; 330 A; 447 C; 424 G; 251 T; 0 other;

Query Match 77.2%; Score 1331; DB 25; Length 1452;
Best Local Similarity 99.4%; Pred. No. 1.7e-294;
Matches 1367; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

QY 219 GCGCCCGGATGCGCTCTGGAGCCGCGATGCCACCA-GGGGCGCCATGGGCCCCC 277
DB 78 GCGCCCGGATGCGCTCTGGAGCCGCGATGCCACCAAGGGGGCGCCATGGGCCCCC 137
QY 278 GGGCTCCCGTACATGGGAGCCCGCGCTGGACCCGCGCTGGCCCCCGGG-ATGGA 336
DB 138 GGGCTCCCGTACATGGGAGCCCGCGCTGGACCCGCGCTGGCCCCCGGGATGGA 197
QY 337 GCGCCCGCGAAGCGAGCAGCGCCCGCGCC-GGAGAGCCAGGACAGAGCCAGGCCA 395
DB 198 GCGCCCGCGAAGCGAGCAGCGCCCGCGCCGCGGAGAGCCAGGACAGAGCCAGGCCA 257
QY 396 GCGGAGCCACCGCCCGCGCGAGCGCGAGTCCAGAGGAGGAGAGATGGCTGACAA 455
DB 258 GCGGTGGCCACCGCCCGCGCGAGCGCGAGTCCAGAGGAGGAGAGATGGCTGACAA 317
QY 456 AATCTCTCCCTCAAAGGATTCGGAGCTGGTCCCGAGTCCCAGGCTTACATGGACCTCT 515
DB 318 AATCTCTCCCTCAAAGGATTCGGAGCTGGTCCCGAGTCCCAGGCTTACATGGACCTCT 377
QY 516 GCATTTGAGAGAACTGGATCAAAACCATCATATCGGAAGGGGTGGACATCCAGAGGC 575
DB 378 GCATTTGAGAGAACTGGATCAAAACCATCATATCGGAAGGGGTGGACATCCAGAGGC 437
QY 576 TCTGAAGAGCCCATGAGCAAAAGCGGAGTGGACCTCTATATCTCCACACTTTAA 635
DB 438 TCTGAAGAGCCCATGAGCAAAAGCGGAGTGGACCTCTATATCTCCACACTTTAA 497
QY 636 CCTCGCAAGTCTGATGTGAGGATTCGACCGGAGCATTTGCTCTCTGGAGCTACGGGT 695
DB 498 CCTCGCAAGTCTGATGTGAGGATTCGACCGGAGCATTTGCTCTCTGGAGCTACGGGT 557
QY 696 GGAGGGGAGCTCTGGATGATCCAGCAAAAGCGGAGGAGTTCTCTCTTTCTTCAA 755

DB 558 GGAGGGGAAGCTCTCGATGATCCAGCAAAACAGAGGGAAGTTCTCTTCTTTCTTCAA 617
QY 756 GAGTTTGGTTCATCGAGCTGGACAAAGATCTTTATGGCCCTGACAAACACCTCGTTGAGTG 815
DB 618 GAGTTTGGTTCATCGAGCTGGACAAAGATCTTTATGGCCCTGACAAACACCTCGTTGAGTG 677
QY 816 GCATCGGACACCCACAGACCCAGAGAGCGGCTTCCAGGTGAAAACGGCTTGGGACCT 875
DB 678 GCATCGGACACCCACAGACCCAGAGAGCGGCTTCCAGGTGAAAACGGCTTGGGACCT 737
QY 876 GAGTGTGGCTGACAGCTGCTCTCATGCTGGACTACAGCTCCCGAGTTCAAACCTGGA 935
DB 738 GAGTGTGGCTGACAGCTGCTCTCATGCTGGACTACAGCTCCCGAGTTCAAACCTGGA 797
QY 936 TCCCGCGCTAGCCCGGCTGCTGGGCTGCACACACAGAGCCGCTCAGCCATTGTCCAGGC 995
DB 798 TCCCGCGCTAGCCCGGCTGCTGGGCTGCACACACAGAGCCGCTCAGCCATTGTCCAGGC 857
QY 996 CCTGTGGAGTATGTGAAGACCAAACAGGCTGAGGACTCCCATGACAAAGGAATACATCAA 1055
DB 858 CCTGTGGAGTATGTGAAGACCAAACAGGCTGAGGACTCCCATGACAAAGGAATACATCAA 917
QY 1056 TGGGGAACAGTATTTCCAGCAGATTTTGAATGTCGCCGCTGAAGTTTCTGAGATTC 1115
DB 918 TGGGGAACAGTATTTCCAGCAGATTTTGAATGTCGCCGCTGAAGTTTCTGAGATTC 977
QY 1116 CCAGCGCTCACAGCCCTGCTTATTCGCCCTGACCCCAATTTGTCATCAACCATGTCATCAG 1175
DB 978 CCAGCGCTCACAGCCCTGCTTATTCGCCCTGACCCCAATTTGTCATCAACCATGTCATCAG 1037
QY 1176 CGTGGACCTTTCAGACAGAAAGAGAGCGCTGCTATGACATTCAGTGGAGGTGGAGGA 1235
DB 1038 CGTGGACCTTTCAGACAGAAAGAGAGCGCTGCTATGACATTCAGTGGAGGTGGAGGA 1097
QY 1236 GCCATTTAAGGGGAGATGAGCAGCTTCTCTCTATCCAGGCCAACCCAGCAGGAGATCAG 1295
DB 1098 GCCATTTAAGGGGAGATGAGCAGCTTCTCTCTATCCAGGCCAACCCAGCAGGAGATCAG 1157
QY 1296 TCCTCTGGACAGTAAAGATCCATGAGAGCATTTGAGTCCATAAACCCAGCTCAAGATCCAGAG 1355
DB 1158 TGCTCTGGACAGTAAAGATCCATGAGAGCATTTGAGTCCATAAACCCAGCTCAAGATCCAGAG 1217
QY 1356 GGAATTCATGCTAAGCTTCTCCAGAGACCCCAAGGCTATGTCCAAAGCTGCTCCGCTC 1415
DB 1218 GGAATTCATGCTAAGCTTCTCCAGAGACCCCAAGGCTATGTCCAAAGCTGCTCCGCTC 1277
QY 1416 CCAGAGCCGGACCTCAAGTGTATGACAGATGTAGCCGGCAACCTGAGAGGAGCGCCG 1475
DB 1278 CCAGAGCCGGACCTCAAGTGTATGACAGATGTAGCCGGCAACCTGAGAGGAGCGCCG 1337
QY 1476 GGCTGAGTTCTACACAGCCCTGCTCCAGGAGGCGGTCAAGTCCGCTACTTCTACTGCAA 1535
DB 1338 GGCTGAGTTCTACACAGCCCTGCTCCAGAGGCGGTCAAGTCCGCTACTTCTACTGCAA 1397
QY 1536 GATCCAGCGCGCAGGAGGAGCTGGAGCAGTCCGTGTTGTGCGCAACACCTAG 1590
DB 1398 GATCCAGCGCGCAGGAGGAGCTGGAGCAGTCCGTGTTGTGCGCAACACCTAG 1452

RESULT 2
AAH98221
ID AAH98221 standard; cDNA; 3089 BP.
XX
AC AAH98221;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 78.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.

```

XX OS Homo sapiens.
XX PN WO200154477-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02687.
XX PR 25-JAN-2000; 2000US-0491404.
XX PR 17-JUL-2000; 2000US-0617746.
XX PR 03-AUG-2000; 2000US-0631451.
XX PR 15-SEP-2000; 2000US-0663870.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Dymnac RA, Zhang J, Wehrman T;
XX DR MPI; 2001-476164/51.
XX DR P-PSDB; AAM23562.
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX PT antibodies and research use -
XX PS Claim 1; Page 231-232; 1275pp; English.
XX CC The present invention provides the protein and coding sequences of novel
XX CC proteins from a variety of organisms, including human, dog, cat, horse,
XX CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX CC from the organism of interest. They can be used in diagnostics,
XX CC forensics, gene mapping, identification of mutations, to assess
XX CC biodiversity and for nutritional purposes. The present sequence is a cDNA
XX CC of the invention.
SQ Sequence 3089 BP; 648 A; 954 C; 946 G; 541 T; 0 other;

Query Match 76.7%; Score 1322.6; DB 22; Length 3089;
Best Local Similarity 91.8%; Pred. No. 1.8e-292;
Matches 1637; Conservative. 0; Mismatches 79; Indels 67; Gaps 20;

QY 1 GAATTCGCGCGAGCGCCGAGCCGAGCGCGAGCGAGCGAGCGAGCGCGCGCTCCGG 60
DB 1 GAATTCGCGCGAGCGCCGAGCCGAGCGCGAGCGAGCGAGCGAGCGCGCGCTCCGG 60
QY 61 GCCGGGGTCCCGGGGAGCAGATCCTCAGAAATGGCCCTTGTGCTGCAAGCGCGTGGGC 120
DB 61 GCCGGGGTCCCGGGGAGCAGATCCTCAGAAATGGCCCTTGTGCTGCAAGCGCGTGGGC 120
QY 121 TCCGGGCGCCGAGCAGCGAGGGGGCACTGATGACTTCTCAGGTGCAAGACCTTGCATCT 180
DB 121 TCCGGGCGCCGAGCAGCGAGGGGGCACTGATGACTTCTCAGGTGCAAGACCTTGCATCT 180
QY 121 TCCGGGCGCCGAGCAGCGAGGGGGCACTGATGACTTCTCAGGTGCAAGACCTTGCATCT 180
DB 121 TCCGGGCGCCGAGCAGCGAGGGGGCACTGATGACTTCTCAGGTGCAAGACCTTGCATCT 180
QY 181 ATGACTCCAGGCTTCAAGCACCACCGTGTACAGCGCCGCGGATGCGTCTGGA 240
DB 181 ATGACTCCAGGCTTCAAGCACCACCGTGTACAGCGCCGCGGATGCGTCTGGA 240
QY 181 ATGACTCCAGGCTTCAAGCACCACCGTGTACAGCGCCGCGGATGCGTCTGGA 240
DB 181 ATGACTCCAGGCTTCAAGCACCACCGTGTACAGCGCCGCGGATGCGTCTGGA 240
QY 241 GCCCGGATGCCCAACA-GGGCGCCCATGAGCGCCCGCGGCTCCCGTATCATGGGCAAC 299
DB 241 GCCCGGATGCCCAACA-GGGCGCCCATGAGCGCCCGCGGCTCCCGTATCATGGGCAAC 299
QY 241 GCCCGGATGCCCAACA-GGGCGCCCATGAGCGCCCGCGGCTCCCGTATCATGGGCAAC 299
DB 241 GCCCGGATGCCCAACA-GGGCGCCCATGAGCGCCCGCGGCTCCCGTATCATGGGCAAC 299
QY 300 CCGCGCGTGCACCGCGGCTTGGCCCGCGGGG-ATGAGACCGCGCCCGCAAGGAGACGCG 358
DB 301 CCGCGCGTGCACCGCGGCTTGGCCCGCGGGG-ATGAGACCGCGCCCGCAAGGAGACGCG 358
QY 301 CCGCGCGTGCACCGCGGCTTGGCCCGCGGGG-ATGAGACCGCGCCCGCAAGGAGACGCG 358
DB 301 CCGCGCGTGCACCGCGGCTTGGCCCGCGGGG-ATGAGACCGCGCCCGCAAGGAGACGCG 358
QY 359 CCGCGCGCC-GGAGAGCCAGGACAGAGCAGGGGCGAGCGGAGCCACCGCCCGCGG 417
DB 361 CCGCGCGCC-GGAGAGCCAGGACAGAGCAGGGGCGAGCGGAGCCACCGCCCGCGG 417
QY 361 CCGCGCGCC-GGAGAGCCAGGACAGAGCAGGGGCGAGCGGAGCCACCGCCCGCGG 417
DB 361 CCGCGCGCC-GGAGAGCCAGGACAGAGCAGGGGCGAGCGGAGCCACCGCCCGCGG 417
QY 418 CGGAGCCGAGTGCAGAG-----GAGGAAGTGGCTGACAA 455
DB 421 CGGAGCCGAGTGCAGAG-----GAGGAAGTGGCTGAGG 480

```

```

QY 456 AATCCTCCCTCAAGAGATTCCGAGCTGTCCCGAGTCCGAGCTTACATGACCTCTT 515
DB 481 GAACCGCAGGAGCAGGATTCGGGAGCTGTCCCGAGTCCGAGCTTACATGAGACTCTT 540
QY 516 GGCATTGAGAGGAACTGATCAACCATATGCGGAGGGGTGGACATCCAGAGGC 575
DB 541 GGCATTGAGAGGAACTGATCAACCATATGCGGAGGGGTGGACATCCAGAGGC 600
QY 576 TCTGAAGAGCCCATGAGCAAAAGCGGAGCTGATCTATATCTTCAACACTTTAA 635
DB 601 TCTGAAGAGCCCATGAGCAAAAGCGGAGCTGATCTATATCTTCAACACTTTAA 660
QY 636 CCCTGCGAAGCTGATGCTGAGATTCCGAGCGGAGCATTCCTCTGGAGCTACGGGT 695
DB 661 CCCTGCGAAGCTGATGCTGAGATTCCGAGCGGAGCATTCCTCTGGAGCTACGGGT 720
QY 696 GGAGGGAGAGCTCCTGATGAT-----CCGAGCAAGAGCGGAAAGT 739
DB 721 GGAGGGAGAGCTCCTGATGATGATCGTCCGCGCCAGCCAGCAAGAGCGGAAAGT 780
QY 740 TCTCTTCTTCTTCAAGGTTTGGTCAATGAGCTGGAACAAAGATCTTATGGCCCTGACA 799
DB 781 TCTCTTCTTCTTCAAGGTTTGGTCAATGAGCTGGAACAAAGATCTTATGGCCCTGACA 840
QY 800 ACCACCTGTTGATGTCATGAGACACCAAGAGCGGAGCGGCTTCCAGTGA 859
DB 841 ACCACCTGTTGATGTCATGAGACACCAAGAGCGGAGCGGCTTCCAGTGA 900
QY 860 AACGGCTGGGAGCCTGATGTCGCTGACAGCTGCTCTCATGCTGAGACTACAGCCTC 919
DB 901 AACGGCTGGGAGCCTGATGTCGCTGACAGCTGCTCTCATGCTGAGACTACAGCCTC 960
QY 920 CCCAGTTAACTGATCCCCCGCTAGCCGCGCTGCTGGGCTGACACACAGAGCCCT 979
DB 961 CCCAGTTAACTGATCCCCCGCTAGCCGCGCTGCTGGGCTGACACACAGAGCCCT 1020
QY 980 CAGCATTTGTCAGGCGCCGTGAGCATGATGTAAGAGCAACAGGCTGAGAGCTCCCATG 1039
DB 1021 CAGCATTTGTCAGGCGCCGTGAGCATGATGTAAGAGCAACAGGCTGAGAGCTCCCATG 1080
QY 1040 ACAAGGAATATCATCATGAGGAGCAAGATTTTCAAGAGATTTTGAATTTCCCGGCTGA 1099
DB 1081 ACAAGGAATATCATCATGAGGAGCAAGATTTTCAAGAGATTTTGAATTTCCCGGCTGA 1140
QY 1100 AGTTTCTGAGATTCCCAAGGCTCACAAGCCCTGCTATTTGCCCTTGAACCAATTGCA 1159
DB 1141 AGTTTCTGAGATTCCCAAGGCTCACAAGCCCTGCTATTTGCCCTTGAACCAATTGCA 1200
QY 1160 TCAACCATGATATCAAGGCTGAGACCTTGA-CCAGAGAGAGACAG-CGTGCTATGACAT 1217
DB 1201 TCAACCATGATATCAAGGCTGAGACCTTGA-CCAGAGAGAGAGAGCGTGCTATGACAT 1260
QY 1218 TGAAGTGAAGTGAAGAGCC--ATTAAAGGGGAGATGAGCAGCTTCT--CCTATCA 1273
DB 1261 TGAAGTGAAGTGAAGAGCCATTAAAGGGGAGAGATGAGCAGCTTCTCTTCTATTTCA 1320
QY 1274 CGGCGCAACAGAGAGATCACT---CCTTGAACATTAAGAT--CAATGAGAGATTGA 1328
DB 1321 CGGCGCAACAGAGAGATCACTCTTCTTGAACATTAAGATCCATGAGAGCATTTGA 1380
QY 1329 GTCCATTAACAGCTCAAG-----TCCAGAGAGGCTTC-ATGCTTAAGCTTCCAGAGA- 1382
DB 1381 GTCCATTAACAGCTCAAGCTCCAGAGATCCAGAGAGGCTTAAAGCTTCTTCCAGAG 1440
QY 1383 --CCCAAGGCTATGTCCAAGACCTGCTCCGCTCCAGAGCGGAGCCTCAAGG--TGA 1438
DB 1441 ACCCCCAAGGCTATGTCCAAGACCTGCTCCGCTCCAGAGCGGAGCCTTCAAGGTTGA 1500
QY 1439 TGAAGATGTAAGCGGCAACCTTGAAGAGAGCGCGGCG--TGAATTCAACACC-AGCC 1496
DB 1501 TGAAGATGTAAGCGGCAACCTTGAAGAGAGCGCGGCGGTTGAATTCAACACCAGCC 1560

```



```

Db      1081 ACAAGAAATACATCATGGGGACAAGTATTTCCAGACATTTTGATTTGCCCGGCTGA 1140
Qy      1100 AGTTTCTAGATTTCCCGAGCGGCTCACAGCCCTGTATTTGCCCGGCTGA 1159
Db      1141 AGTTTCTAGATTTCCCGAGCGGCTCACAGCCCTGTATTTGCCCGGCTGA 1200
Qy      1160 TCACCATGTCTACAGCTGAGACCTTTCAGA--CCAGAGAAGACAG--CGTGTATGACAT 1217
Db      1201 TCACCATGTCTACAGCTGAGACCTTTCAGACCCAGAGAAGCGGTGTGTATGACAT 1260
Qy      1218 TCACGTGAGGTGAGAGAGCC--ATTAAAGGGCAGATGAGAGCTTCTT--CTATGCA 1273
Db      1261 TCACGTGAGGTGAGAGAGCCCATTAAGGGGCGCAGATGAGAGCTTCTTCTATTTCA 1320
Qy      1274 CGGCGCACAGCAGAGGATCAGT---CCTTGAGACATGAT--CATTGAGACATTTGA 1328
Db      1321 CGGCGCAACAGCAGAGAGATCAGTCTTCTGAGACATGATCCCATGAGCCGATTTGA 1380
Qy      1329 GTCCATTAACCAAGCTCAGA---TCAGAGGGAGCTTC--ATGCTAGCTTCTTCAGAGA- 1382
Db      1381 GTCCATTAACCAAGCTCAGAAGATCCAGAGGAGCTTCAATGCTTAAGTTTCTTCAGAG 1440
Qy      1383 --CCCCAAAGGCTATGTCCAGAAGCTGCTCCGCTCCAGAGCGGAGCCTCAAG--TGA 1438
Db      1441 ACCCCCAAGGCTATGTCCAGAAGCTGCTCCGCTCCAGAGCGGAGCCTCAAGGTTGA 1500
Qy      1439 TCACAGATATAGCCGCGCAACCTTGAAAGAGAGGCGCGGCG--TGAGTTTCAACAC--AGCC 1496
Db      1501 TCACAGATATAGCCGCGCAACCTTGAAAGAGAGGCGCGGCTTGAGTTTCAACACAGCC 1560
Qy      1497 CCGGTCCAGAGAGCGGCTCAGTC--GGTACTTCTAC--TGCAAGATCCAGAGCGCAGAG 1554
Db      1561 CTGGTCCAGAGAGCGGCTCAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1620
Qy      1555 GAGCTGAGACAGTCTGCTGTTTGCCCAACACTTAGAGAGCCCAAAACAGACAGCAGC 1614
Db      1621 GAGCTGAGACAGTCTGCTGTTTGCCCAACACTTAGAGAGCCCAAAATAGACAGCAGC 1680
Qy      1615 GGAACCTTTCAGCGGCTTCCCGGCGCCAGCATTTCCTCCGCGGCTCCAGC--TCATCTTC 1673
Db      1681 GGAACCTTTCAGCGGCTTCCCGGCGCCAGCATTTCCTCCGCGGCTCCAGCATTCTTC 1740
Qy      1674 TCGCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCATCTG 1716
Db      1741 TCGCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCATCTG 1783

RESULT 4
AB054477
ID      AB054477 standard; cDNA, 1970 BP.
XX
XX
AC      AB054477;
XX
XX
DT      22-AUG-2002 (first entry)
XX
XX
DE      Human ovarian antigen HSD108 cDNA, SEQ ID NO:357.
XX
XX
KW      Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW      ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW      infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW      PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW      inflammatory condition; immune disorder; blood disorder;
KW      cardiovascular disorder; respiratory disorder; neurological disorder;
KW      gastrointestinal disorder; urinary system disorder; drug screening;
KW      gene therapy; chromosome mapping; forensic analysis;
KW      antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW      antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS      Homo sapiens.
XX
XX      WO200200677-A1.
XX

```

```

PD      03-JAN-2002.
XX
XX      07-JUN-2001; 2001WO-US18569.
XX
XX      07-JUN-2000; 2000US-209467P.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Birse CE, Rosen CA;
XX
XX      WPI, 2002-147878/19.
XX
XX      P-PSDB; ABP41400.
XX
XX      Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX      useful in the prevention, treatment and diagnosis of cancer (e.g.
XX      ovarian cancer), immune disorders, cardiovascular disorders and
XX      neurological diseases -
XX
XX      Claim 1; SEQ ID No 357; 2922bp; English.
XX
XX      The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX      ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
XX      encompasses polypeptides 90% identical and polynucleotides 95% identical
XX      to the sequences of the invention. The invention additionally relates to
XX      recombinant vectors and host cells comprising human ovarian antigen
XX      polynucleotides, antibodies against human ovarian antigen, and the use
XX      of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX      treating, prognosting or preventing various ovary and/or breast-related
XX      disorders. Such conditions include ovarian cancer and breast cancer, and
XX      metastatic tumours of ovarian or breast origin, reproductive system
XX      disorders (e.g., infertility, disorders of pregnancy, anovulation,
XX      polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
XX      disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
XX      shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
XX      vaginitis), immune disorders (e.g., congenital and acquired
XX      immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
XX      blood-related disorders (e.g., anaemia), cardiovascular disorders,
XX      respiratory disorders, neurological disorders, gastrointestinal disorders
XX      and urinary system disorders. Ovarian antigen polypeptides and
XX      polynucleotides may also be used in screening for compounds which
XX      modulate ovarian antigen expression or activity. The polynucleotides may
XX      further be used for gene therapy, chromosome mapping, in the
XX      identification of individuals and in forensic analysis, and the
XX      polypeptides may be used as food additives or to prepare antibodies
XX      useful in disease diagnosis, drug targeting and phenotyping. The present
XX      sequence represents cDNA encoding a human ovarian antigen of the
XX      invention.
XX
XX      CC Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      SQ      Sequence 1970 BP; 463 A; 590 C; 522 G; 389 T; 6 other;

Query Match      57.0%; Score 983; DB 24; Length 1970;
Best Local Similarity 99.0%; Pred. No. 6; 7e-215;
Matches 1027; Conservative 3; Mismatches 3; Indels 4; Gaps 4;

```

QY 285 CCGTACATGGGAGCCCCCGCGTGCACCCGGCTTGGCCCCCGCGG-ATGAGAGCCCC 343
Db 241 CCGTACATGGGAGCCCCCGCGTGCACCCGGCTTGGCCCCCGCGG-ATGAGAGCCCC 300
QY 344 CCGAAGCGAGAGCGCCCCCGCGG-GGCAGAGCCAGGACACAGAGCCAGGCGCGGAG 402
Db 301 CCGAAGCGAGAGCGCCCCCGCGG-GGCAGAGCCAGGACACAGAGCCAGGCGCGGAG 360
QY 403 CCCACGCGCCCGCGGAGCGCGAGTCCAAAGAGGAGGAAGTGGCTGACAAAATCTCTC 462
Db 361 CCCACGCGCCCGCGGAGCGCGAGTCCAAAGAGGAGGAAGTGGCTGACAAAATCTCTC 420
QY 463 COTCAAAGGATTCGGAGCTGGTCCCGAGTCCAGGCTTACATGAGACCTCTTGGCAATT 522
Db 421 COTCAAAGGATTCGGAGCTGGTCCCGAGTCCAGGCTTACATGAGACCTCTTGGCAATT 480
QY 523 GAGAGAACTGGATCAAAACCATCATGCGGAAGGGGGTGGACATCCAGGAGCTCTGAAG 582
Db 481 GAGAGAACTGGATCAAAACCATCATGCGGAAGGGGGTGGACATCCAGGAGCTCTGAAG 540
QY 583 AGSCCATGAAGCAAAAGCGGAGCTGGAGCTTATATCTCCAAACACTTTTAAACCTCGG 642
Db 541 AGSCCATGAAGCAAAAGCGGAGCTGGAGCTTATATCTCCAAACACTTTTAAACCTCGG 600
QY 643 AAGTCTGATGCTGAGGATTCGACGCGCAGCATTTGCTCTCTGGAGCTACGGGTGAGGG 702
Db 601 AAG-CTGATGCTGAGGATTCGACGCGCAGCATTTGCTCTCTGGAGCTACGGGTGAGGG 659
QY 703 AAGTCTGATGATCCAGCAAAACAGAAAGCGGAAGTTCTCTTCTTCTTCAAGATTG 762
Db 660 AAGTCTGATGATCCAGCAAAACAGAAAGCGGAAGTTCTCTTCTTCTTCAAGATTG 719
QY 763 GTCATCGAGCTGCAAAAGATCTTTATGSCCTGTGACACACCTCGTTGAGTGGCATCGG 822
Db 720 GTCATCGAGCTGCAAAAGATCTTTATGSCCTGTGACACACCTCGTTGAGTGGCATCGG 779
QY 823 ACACCCAGCAGCAGGAGACGCGCTTCCAGGTGAAACGCGCTGGGACCTGAGTGTG 882
Db 780 ACACCCAGCAGCAGGAGACGCGCTTCCAGGTGAAACGCGCTGGGACCTGAGTGTG 839
QY 883 CGCTGACGCTGCTCTCATGTGTGACATACAGCTTCCAGTTCAGGTTCAAACTGATCCCGC 942
Db 840 CGCTGACGCTGCTCTCATGTGTGACATACAGCTTCCAGTTCAGGTTCAAACTGATCCCGC 899
QY 943 CTAGCCGCGCTGCTGGGCTGCACACAGAGCGCTCAGCATTTGTCAGGSCCTGTGG 1002
Db 900 CTAGCCGCGCTGCTGGGCTGCACACAGAGCGCTCAGCATTTGTCAGGSCCTGTGG 959
QY 1003 CAGTATGTGAAGACCAACAGGCTGAGGACTCCCATGACAGGAATACATCAATGGGAC 1062
Db 960 CAGYATGKAAGACCAACAGGCTGAGGACTCCCATGACAGGAATACATCAATGGGAC 1019
QY 1063 AAGTATTCACAGAT 1079
Db 1020 AAGTATTCACAGAT 1036

RESULT 5

ABZ35936/c
ID ABZ35936 standard; cDNA; 6809 BP.

XX

AC ABZ35936;

XX

DT 10-FEB-2003 (first entry)

XX

DE Human secretory polynucleotide SPTM SEQ ID NO 100.

XX

KW Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;
KW anti-inflammatory; immunosuppressive; neuroprotective; neurotropic;
KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;

KW

KW

XX

OS Homo sapiens.

XX

PN WO200283876-A2.

XX

PD 24-OCT-2002.

XX

XX 27-MAR-2002; 2002WO-US09921.

XX

PR 29-MAR-2001; 2001US-280067P.

PR

PR 29-MAR-2001; 2001US-280068P.

PR

PR 16-MAY-2001; 2001US-291280P.

PR

PR 17-MAY-2001; 2001US-291829P.

PR

PR 17-MAY-2001; 2001US-291849P.

PR

PR 19-JUN-2001; 2001US-299428P.

PR

PR 20-JUN-2001; 2001US-299776P.

PR

PR 20-JUN-2001; 2001US-300001P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;

PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta C, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RV, Urashka ME;

XX

XX WPI; 2003-075543/07.

DR P-PSDB; ABP75489.

DR

XX

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Claim 1; SEQ ID NO 100; 458pp + Sequence Listing; English.

The invention relates to a secretory polynucleotide (designated spm) comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a naturally occurring polynucleotide sequence at least 90 % identical to the polynucleotide sequence, a polynucleotide complementary to them or an RNA equivalent of them. The polypeptide or polynucleotide are useful for treating, preventing or diagnosing a disease or condition associated with the expression of functional SPTM. These are particularly useful for diagnosing, treating or preventing autoimmune/inflammatory disorders (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's disease), neurological disorders (e.g. epilepsy, Huntington's disease, dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease, multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety, schizophrenia or amnesia), or cell proliferative disorders (e.g. psoriasis, polycythemia vera, or cancers including adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain, breast, cervix or prostate).
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 6809 BP; 1248 A; 1988 C; 2042 G; 1530 T; 1 other;

Query Match 47.7%; Score 821.6; DB 25; Length 6809;

Best Local Similarity 99.2%; Pred. No. 7.5e-178;

Matches 857; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 219 GCGCCCGGGATGCGCTGTGGAGCCCGGATGCCCCACCA-GGGGCGCCATGGGCCCCCC 277

Db 6712 GCGCCCGGGATGCGCTGTGGAGCCCGGATGCCCCACCA-GGGGCGCCATGGGCCCCCC 6653

QY 278 GGGTCTCCCGTACATGGGACCCCGCGTGCAGCCCGCTGCCCCCGGG-ATGGA 336

Db 6652 GGGTCTCCCGTACATGGGACCCCGCGTGCAGCCCGCTGCCCCCGGGCATGGA 6593

QY 337 GCCCGCCCGCAAGCAGCAGCGCCCCGCC-GGCAGAGCCAGGCACAGAGCCAGGGCCA 395

QY 735 GAAGTTCCTCTTTCTTCAAGAGTTTGGTCAATCGAGCTGGCAAAAGATCTTTATGGCCC 794
Db 684 AAGGTTCTCTCTTTTAAAGTCTTGGTATCGAATGGCAAAAGACCTCTATGGCCC 743
QY 795 TGACAAACACCTCGTTGAGTGCATCGACACCCACGACCCAGGAGACGGCGCTTCCA 854
Db 744 AGACAAACCATCTGGTAGAATGCAAGGACCGCCACTACCCAGGAGACGATGGCTTCCA 803
QY 855 GGTGAAACGGCTGGGGACCTCAGTGTGGCTGACGCTGCCTCATCTGCTGGACTACCA 914
Db 804 GGTGAAGCGCCAGAGATGTGAATGTACGGTGTACTGTCTGCTGATGCTGGACTACCA 863
QY 915 GCCTCCCAAGTTCAAACTCGATCCCGCTAGCCCGCTGCTGGGGCTGCAACACAGAG 974
Db 864 GCGCCGCCAGTTTAAATTTAGACCTCGCTCGCTCTGGGCTATCCATACCCAGAC 923
QY 975 CCGCTCAGGCATGTTCAGGCCCTGTGGCAGTATGTGAAGCAACACAGCTGCAGGATC 1034
Db 924 ACGTCCAGTGATCATCCAAGCACTGTGGCAGTATATTAACACACAAAGCTCCAGGACCC 983
QY 1035 CCATGACAAGGAATACATCAATGGGGACAAGTATTTCCAGCAGATTTTGTGATTTGCCCG 1094
Db 984 TCACGAGCAGAGTTTGTCTCTGTGACAGTACCTCCAGCAGATCTTTGATCTCAGGG 1043
QY 1095 GCTGAAGTTTCTGAGATTTCCCAAGCGCTCAAGCCCTGTCTATTTGCCCGCTGACCCCAAT 1154
Db 1044 GATGAAGTTCTCAGAGATCCCTCAGCGGCTCCACGCTTGCTTATGCCACAGAGCCCAT 1103
QY 1155 TGTATCAACATGTCATCAGCGTGGACCCCTTCAGACCAGAGACAGCGTGTATGA 1214
Db 1104 CATCATCAATCATGTCTCAGTGTGGACCCCAATGACCAGAAAAAGACCGGTGTATGA 1163
QY 1215 CATTGACGTGGAGGTGGAGGCCATTAAAGGGGAGATGACGACTTCTCTCTATTCAC 1274
Db 1164 CATTGACGTGGAGGTGGATGACACTCTGAAGCCAGATGAATCTTCTCTGTGTTCCAC 1223
QY 1275 GCGCAACACGACGAGAGATCAGTCTCTCGACAGTAAAGATCCATGAGACGATTTAGTCCAT 1334
Db 1224 TGCCAGCCAGCAGGAGATCGCCACTCTAGACAAAGATCCATGAGCAGATAGACACCAT 1283
QY 1335 AAACAGCTCAAGATCCAGAGGACTTCAATGCTTAAGCTTCTCCAGACCCCAAGGCTA 1394
Db 1284 CAACAGCTTGAGACCCAGCAGAGTTCAATGTTGAGCTTGCCC---GAGCCCTCAGGTTT 1340
QY 1395 TGTCAAGACCTGCTCCGCTCCAGAGCGGACCTCAAGGTGATGACAGATGTAGCCGG 1454
Db 1341 CATCAATGATTGGCTTCAGTCCAGTCGAGGGCCCTCAAGCAGTACTGATGTGGTGG 1400
QY 1455 CAACCTGAAGAGGAGCGCGGCTGAGTTCTTACCAACAGCCCTCGTCCAGGAGCGCT 1514
Db 1401 TAAACCGGAAGAGGAGCGTGTGCTGAGTTCTTCTTCCAGCCCTGGGCTCAGGAGGCTGT 1460
QY 1515 CAGTCGCTACTTCTACTGCAAGATCCAGAGCGCAGGAGGCTGGAGCAGTGCCTGGT 1574
Db 1461 GTGCGGATATCTTACTTCAAGGTGCAGCAGAGGGCGCAAGAGTTAGAGCAAGCCCTGGG 1520
QY 1575 TGTGCGCAACACTAGGAGCCC 1596
Db 1521 AATCCGAACATAGAGGCTC 1542

RESULT 7
ABS73872
ID ABS73872 standard; cDNA; 2511 BP.
XX AC ABS73872;
XX DT 06-DEC-2002 (first entry)
XX DE Human cDNA encoding NAAPl2, Incyte 3112390C81.
XX Human; ss; gene; nucleic acid-associated protein; NAAp; gene therapy;
KW microarray; proliferative disorder; developmental disorder;

cardiovascular disorder; neurological disorder; autoimmune disorder;
inflammatory disorder; atherosclerosis; bursitis; cirrhosis; hepatitis;
psoriasis; cancer; lymphoma; melanoma; brain cancer; breast cancer;
Cushing's syndrome; Alzheimer's disease; AIDS; Creutzfeldt-Jakob disease;
acquired immunodeficiency syndrome; Goodpasture's syndrome;
Crohn's disease; multiple sclerosis.
Homo sapiens.
W0200274913-A2.
26-SEP-2002.
14-MAR-2002; 2002WO-US07869.
16-MAR-2001; 2001US-276857P.
19-APR-2001; 2001US-285489P.
19-APR-2001; 2001US-285558P.
04-MAY-2001; 2001US-288646P.
04-MAY-2001; 2001US-288700P.
10-MAY-2001; 2001US-290510P.
11-MAY-2001; 2001US-290369P.
16-NOV-2001; 2001US-332426P.
(INCY-) INCYTE GENOMICS INC.
Hillman JL, Baughn MR, Swarnakar A, Yue H, Elliott VS, Burford N;
Ding L, Tang YT, Lee SY, Azimzai Y, Wallia NK, Gietzen KJ;
Griffin JA, Lai PG, Yang J, Borowsky ML, Richardson TW, Yue H;
Becha S, Forsythe IJ, Jones KA, Warren BA, Thangavelu K;
Honchell CD, Jolley HE, Hafalia AJA, Ring HZ;
WPI; 2002-723448/78.
p-PSDB; ABG95655.
New human nucleic acid-associated proteins and polynucleotides, useful
for diagnosing, treating or preventing cardiovascular, neurological,
autoimmune or inflammatory disorders, e.g. atherosclerosis, Alzheimer's
or AIDS -
Claim 5; Page 205-206; 219pp; English.
The invention relates to an isolated human nucleic acid-associated
protein (NAAp), comprising the protein sequences appearing as ABG95644-
ABG95669, or an immunogenic or biologically active fragment. Also
included are the NAAp encoding polynucleotides, a recombinant
polynucleotide comprising a promoter sequence operably linked to the
NAAp polynucleotide, a cell transformed with the recombinant NAAp
polynucleotide, an NAAp transgenic organism, an anti-NAAp antibody,
screening compounds as modulators or ant/agonists of NAAp or
modulators of NAAp polynucleotide expression, a microarray comprising
30-60 nucleotides of the NAAp polynucleotides and generating an
expression profile of a sample that contains NAAp polynucleotides.
The polypeptide, polynucleotide and agonist are useful for
treating a disease or condition associated with decreased expression of
functional NAAp. The antagonist is useful for treating a disease or
condition associated with overexpression of functional NAAp. The
antibody that specifically binds to the polypeptide is useful for
diagnosing a condition or disease associated with the expression of
NAAp. These diseases include proliferative, developmental,
cardiovascular, neurological, or autoimmune or inflammatory disorders.
These polypeptides, polynucleotides, agonists and antagonists may also
be used for preventing these diseases. These disease or conditions
associated with functional NAAp expression also includes
atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis, cancers
(e.g. lymphoma, melanoma, brain cancer or breast cancer), Cushing's
syndrome, Alzheimer's disease, acquired immunodeficiency syndrome (AIDS),
Creutzfeldt-Jakob disease, Goodpasture's syndrome, Crohn's disease, or
multiple sclerosis and many other diseases and conditions listed in
the specification. The present sequence encodes an NAAp protein.
Sequence 2511 BP; 512 A; 713 C; 760 G; 526 T; 0 other;

Query Match 37.2%; Score 642; DB 24; Length 2511;
 Best Local Similarity 71.1%; Pred. No. 6.4e-137;
 Matches 880; Conservative 0; Mismatches 325; Indels 33; Gaps 1;

QY 393 CCAGCCGAGCCACCCGCGCGGAGCCGAGTGCACAGAGAGAGATGGCTGA 452
 DB 445 CCAGGCGGAGCCCTCCATGCTGCTCCAGCGCGCGGGGTTAAAGAGAGAGATGGCACA 504
 QY 453 CAAATCCCTCCCTCAAAAGATTGGAGGTGGTCCCGAGTCCAGGCTTACATGAGACT 512
 DB 505 TAAAGTTTACCTCAGCGAATCCGGAGCTTTGTTCCAGATCTCAGGCGTACATGATCT 564
 QY 513 CTTCGATTGAGAGAGAACTGATCAACATCATGCGGAGGGGAGTCCAGCA 572
 DB 565 CTTCGCTTTTGAAGCGGAAGCTGAGCCAGACATGCTCCGAGCGGATGGATCCAGGA 624
 QY 573 GGCTTGAAGAGCCCATGAAGCAAAAGCGGAAGCTGCGACTATATCTCAACACTTT 632
 DB 625 GGCATCAAAAAGCCCTCTGACACAAAAGCGAAAGCTTCGATCTACATTTCCAAATGCTT 684
 QY 633 TAACTCGGAAGTCTGATGCTGAGAT----- 660
 DB 685 CAGTCCAGCAAGCGGAAGGCGATAGTGCAGGAAGTGCAGGAGCCCTGGGGGAAACCC 744
 QY 661 -TCCGAGCGAGCATTTGCTCTGAGGAGTACGAGGTGAGAGGAACTCTCGATGATCC 719
 DB 745 AGAGAGGAGACAGAGTGGCTTCTGGGAATCCGAGTGGAGAGAAACTGCTGATATCC 804
 QY 720 CAGCAAAAGAAAGCGGAAGTCTCTTTCTTTCAAGATTTTGTCATGAGCTGACAA 779
 DB 805 TAGCAAAAGAAAGAGAGATTTCTTCATCTTTAAGAGCTGTCATTTAGAGCTGACAA 864
 QY 780 AGATCTTTATGCTGACACCACTCTGTTGATGAGTGCATGCAACCCACAGCCAGGA 839
 DB 865 GAGAGCTGACGAGCTGACATCATCTGTGAGTGCACCCGATCCACACCCAGGA 924
 QY 840 GACGAGCGGCTTCAGAGTGAAGCGGCTGGAGCTGAGTGTGCGTGCAGCGTCTCT 899
 DB 925 GACAGATGGCTTCCAGTAAAGCGGCTGGAGCTCAGACCTCAAGTCAAGCTCTCTGCT 984
 QY 900 CATGCTGACATACAGGCTCTCCAGTTCAACTGATTCGCCGCTTACCGGCTGCTGG 959
 DB 985 CATGCTGATCATCACTCTCCAGTACAAATTTGAGCCCCGATTTGGCAAGGCTGTGG 1044
 QY 960 GCTGCACACAGAGCGGCTCAAGCTTTGTCAGGCGCTGTGCAATGTGAACCA 1019
 DB 1045 AGTGCACAGCAGACGAGGCGGCTCATGAGGCGCTGTGCTTTACATCAAGCA 1104
 QY 1020 CAGGCTGACGAGCTCCATGACAAAGATACATCAATGGGAGCAAGTATTTCCAGCAGAT 1079
 DB 1105 CAGCTGACAGATGGGACGAGGAGGATACATCACTGACCCGTTACTTCCGCCAGAT 1164
 QY 1080 TTTTGAATGCTCCGCTGAAGTTTCTGAGATTTCCCAAGCGCTTCACAGCCCTGCTATT 1139
 DB 1165 CTTCATTTGGGCGGAGCTCCGTTTCTCCAGATTTCCATGAGAGCTGGAGGTTGCTGA 1224
 QY 1140 GCGCCCTGACCAATTTGATCAACCATGTATCAGCTGAGACCTTTCAGACCAAGAA 1199
 DB 1225 GCATCCAGACCCCATTTGATCAACCATGTATTAATGTGACCTTAAGCAGAGAA 1284
 QY 1200 GACAGGCTGATGATGATGAGTGGAGTGGAGGAGCCATTAAAGGGGCAATGAGCAG 1259
 DB 1285 GACAGGCTGATGATGATGAGTGGAGTGGAGGAGCCCACTGAAGGCGCAATGAGCA 1344
 QY 1260 CTTCCTCTATCCAGCGCAACAGAGAGATCACTCTCTGAGACATGAATCCATGA 1319
 DB 1345 TTTTCTGGGCTCTACACCAATCAGAGAGATCGCTCTCTGATGATCAAGTCCATGA 1404
 QY 1320 GACGATTTGATTTAAACAGCTCAAGATCCAGAGGAGCTTATGCTTAAGCTTCTCAG 1379
 DB 1405 GACCATTTGATTTCAACAGCTGAAGCCAGAGAGATTTCTATCTCAAGTTTTCAGC 1464
 QY 1380 AGACCCCAAGAGTATGTCAAGACTGCTCGCTCCAGAGCGGAGACTCAAGGTAT 1439

DB 1465 CACACCCGAGCACTTCATCAGAGATGCTCCGTTCCAGCGCCGAGACTTCAGATCAT 1524
 QY 1440 GACAGATGAGCGGCAACCCCTGAGAGAGAGCGCGGCTGAGTCTACACAGCCCTG 1499
 DB 1525 CACTATGATTTGAAATCTTGAAGAGAGAGAGAGACTGTTCTTACACAGCCCTG 1584
 QY 1500 GTCCAGAGAGCGGCTGATGCTGCTTACTTCAAGATTCAGAGCGCAGAGAGACT 1559
 DB 1585 GCGCCAG 1644
 QY 1560 GAGCAGTCTGCTGTTGCGGCAACACTAGAGCCCA 1597
 DB 1645 GGAACAGGTGCTGGGAATTCGCTGACCTAATGCTCA 1682

RESULT 8
 ABL02583
 ID ABL02583 standard; cDNA; 2010 BP.
 XX
 AC ABL02583;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2231.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI. 2001-658660/75.
 XX
 DR P-PSDB; ABB58480.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 2231; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB16176-AB16175) and the encoded proteins
 CC (AB16173-AB162072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 2010 BP; 509 A; 550 C; 545 G; 406 T; 0 other;

Query Match 33.2%; Score 572.6; DB 23; Length 2010;
 Best Local Similarity 69.9%; Pred. No. 4.4e-121;
 Matches 824; Conservative 0; Mismatches 334; Indels 21; Gaps 3;

QY 433 AAGAGAGAGATGCTGACAAATCTCCCTCAAGATTTGGAGCTGCTCCCGAG 492

Db 581 AAGAAAAAGAGTCGCCGCAAAAGATACTGCCGAGAAGGTGCGGGATCTCGTCCCGGAA 640
Qy 493 TCCAGGCTTACATGCACTCTTGCCATTTGAGAGGAAACTGGATCAAAACCATCATGCGG 552
Db 641 TCGAAGCGTACATGGAATCTGCTGACGTTTGAGCGAANAACCTGGATGCCACCATATGCGC 700
Qy 553 AAGGGGTGGACATCAGAGGCTCTGAAGAGGCCCATGAAGCAAAAGCGGAAGCTGGCA 612
Db 701 AAACGCTGGACATCAGAGGCCCTCAAGCGGCCCATGAAGCAAGCGCAAGCTGCGC 760
Qy 613 CTCTATATCTCAACACATTTAAACCTCGGAGTCTGATGCTGAGGATTC---GAGCGC 669
Db 761 ATCTTATCTCGAACACATTTATCCAGCAGAGGCCCAAAATGATGGCGAGAGGGC 820
Qy 670 AGCATTTGCTCTCGGAGTACGGGTGGAGGGGAAGCTCTCGGATGATCCC----- 720
Db 821 GCCGTGCTCTTGGGAATTCGGGTGGAGGTCCCTGTTAGAGGATGGCAAGGGCGAT 880
Qy 721 -----AGCAAAACAGAAGCGGAAGTTCTTCTTTCTTCAAGAGTTTGGTTCATCGAGCTG 774
Db 881 CCCAATCAAAAGATCAAAACGCAAGTTTTCGTCTGTTTTTCAAGTCGCTGTTTATCGAGCTG 940
Qy 775 GACAAGATCTTTATGGCCCTGACAAACCACTCTGTGAGTGGCATCGACACCCACGACC 834
Db 941 GATAAGSAACTGTACGGTCCGAGCAACCATCTGGTTCGAGTGGCATCGCACTCAACACCA 1000
Qy 835 CAGGAGCGGACGGCTTCCAGGTGAAAACGGCTGGGGACCTGAGTGTGGCTGACACGCTG 894
Db 1001 CAGGAGCGGACGGTTTCAGGTGAAGCGCGCGCGATCGCATGTGCGCTGACCAATC 1060
Qy 895 CTCCTCATCTGGACTACAGCCTCCCGAGTTCAAACCTGGATCCCGCTAGCCCGGCTG 954
Db 1061 CTCCTGCTGCTGACTACAGCGCTGCAGTTCAAGCTGGACCCCGCACTCGCCAGTTG 1120
Qy 955 CTGGGCTGCACACACAGAGCGCTCAGCCATTGTTCAGGCGCCTGTGCGCAGTATGTGAAG 1014
Db 1121 CTGGCGGTACACACACAGACAGGCGGTGTATCATATCCGCCCTGTGGCAGTACATCAAG 1180
Qy 1015 ACCAACAGCTGCAGGACTCCCATCAAGAAATACATCAATGGGGCAAGTATTTCAG 1074
Db 1181 AGCACAAGCTACAGGATGCCCAAGCGGGAGTACATCAATTTGGCAAGATATCTGGAG 1240
Qy 1075 CAGATTTTGTATGTCCTCGGCTGAAGTTTCTGAGATTCCCGAGCCTCACAGCCCTG 1134
Db 1241 CAGATATTCAAGTTGTCAGCGGATGAAGTTTTCAGAGATACCGCAACGCTCAATCCGCTG 1300
Qy 1135 CTATTGCCCCCTGACCAATTTGTATCAACATGTATCATCAGGTGGACCTTCAGACCAAG 1194
Db 1301 CTGCATCCGCCGATCCGATTTGTATCAATCATTTTCAT---TGAGAGCGGTGCAGAGAAC 1357
Qy 1195 AAGAAGCAGCGTGTATGACATTGACGTGGAGTGGAGGACCATTAAGGGGCGAGATG 1254
Db 1358 AAGCAGACTGCTGCTACACATCATGTGAGGTGGATGATACGCTCAAAAACAGATG 1417
Qy 1255 AGCAGCTTCTCTTATCCACGCCCAACAGCAGGAGATCAGTCTCTCGGACAGTAAGATC 1314
Db 1418 AACAGCTTCTCTGATGAGCACTCGAGCCAAACAGGAGATCCAGGGGCTGGACACCAAGATC 1477
Qy 1315 CATGAGACGATGAGTCCATAAACAGCTCAAGATCCAGAGGAGTTTCATGCTAAGCTTC 1374
Db 1478 CACGAGCGGTGGACACGATCAACAGATGAAGACGAAACAGGGAGTTTCTTCTAAGCTTC 1537
Qy 1375 TCCAGAGACCCAAAGGCTATGTCGAAGACCTGCTCCGCTCCAGAGCGGAGCTCAAG 1434
Db 1538 GCCAAGGATCCAAATGTTTATTTCATCTGCGGATCATCAGTGACAGAGGATTTTGAAG 1597
Qy 1435 GTGATGACAGATGTAGCCGCAACCTTGAAGAGAGCGCGGGCTGAGTTCTTACCAACAG 1494
Db 1598 CTGATGACGATGTAGTGGCAATCCGGAGGAGGAGCGTCGGCGGAGTTCTTATTACCAG 1657
Qy 1495 CCCTGGTCCAGGAGGCGTGTAGTCTTCTACTCTCAAGATCCAGAGCGGAGGCGAG 1554
Db 1658 CCATGGACGCGAGGCGGCTTTTCGCGCTACTTTTTCACCAAGGTTCAACCAAGAGCGGCGC 1717

Oy 1555 GAGCTGGAGCAGTCGCTGTTGTGCGCAACACCTAGGAG 1593
Db 1718 GAATTGGAGCAGCGCTGGGCATACGCAACGGCTAGGTG 1756

RESULT 9

ABX34564
ID ABX34564 standard; cDNA; 2581 BP.

XX AC ABX34564;

XX DT 13-FEB-2003 (first entry)

XX DE Human mdt cDNA SEQ ID 125.

XX KW MDDT; human; disease detection and treatment molecule polypeptide;
anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
gene therapy; protein replacement therapy; cell proliferative disorder;
cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;
anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
psoriasis; hepatitis; gene; ss.

XX OS Homo sapiens.

XX PN WO200279449-A2.

XX PD 10-OCT-2002.

XX XX 27-MAR-2002; 2002WO-US09944.

XX XX 28-MAR-2001; 2001US-279619P.

XX XX 29-MAR-2001; 2001US-280067P.

XX XX 29-MAR-2001; 2001US-280068P.

XX XX 16-MAY-2001; 2001US-291280P.

XX XX 17-MAY-2001; 2001US-291829P.

XX XX 17-MAY-2001; 2001US-291849P.

XX XX 19-JUN-2001; 2001US-299428P.

XX XX 20-JUN-2001; 2001US-299776P.

XX XX 20-JUN-2001; 2001US-300001P.

XX XX (INCY-) INCYTE GENOMICS INC.

XX XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

XX XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;

XX XX Daugherty SC, Dam TC, Liu IF, Nguyen DA, Kieseheid Y, Gerstin EH;

XX XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

XX XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX XX WPI: 2003-058431/05.

XX XX P-PSDB; ABU11574.

XX XX Claim 1; SEQ ID NO 125; 339pp + Sequence Listing; English.

XX XX This invention describes a novel disease detection and treatment molecule
polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
and the polypeptides of the invention can be used for gene therapy,
protein replacement therapy and are useful for treating a variety of
diseases or conditions. These polypeptides or polynucleotides are
particularly useful for diagnosing, treating or preventing cell
proliferative disorders (e.g. cancers including adenocarcinoma
leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or

CC hepatitis. ABX34440-ABX34835 encode the MDT polypeptides represented in
 CC ABU11450-ABU11945, described in the disclosure of the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2581 BP; 572 A; 652 C; 737 G; 620 T; 0 other;

Query Match 29.7%; Score 511.4; DB 25; Length 2581;

Best Local Similarity 72.1%; Pred. No. 4,6e-107; Indels 21; Gaps 1;

Matches 690; Conservative 0; Mismatches 246; Indels 21; Gaps 1;

662 CCGACGCGACGATTCCTCTGCGGAGCTACGCGGTGGAGGGAAGTCTCGATGATCCCA 721
 773 CAGGGGACAGAGGCGTCTCTGGAACTCGAATGGAAGAAACGCTCGATGATCCCA 832
 722 GCAACAGAGCGGAAGTCTCTCTCTTCTTCAAGATTGTCATCGAGCTGGACAAG 781
 833 GCAACAGAGGAGGATTTCTTCAATCTTAAAGCCCTGTCATGAGCTGGACAAG 892
 782 ATCTTATGGCCCTGACACACACTGTTG-----AGTGGCATC 820
 893 AGCTGTACGGGCTGACCAATCACTGTGAGCATCTTCCCTCCCTGCACTGGCACC 952
 821 GGAACACCAAGACCCAGAGAGGAGGAGGCTTCCAGGTGAAGCGCTGGGAGCTGAGT 880
 953 GAGTGGCCACACCCAGAGAGAGAGATGGCTTCCAGTAAACCGCTGGAGACTTCAAG 1012
 881 TGGCTGACGCTGCTCTCATGCTGGAATCCAGGCTCCCAATTCAATGATCCCG 940
 1013 TCMAAGTGAACCTCTGCTCATGCTGATCATGACCTCCCGAGTCAAAATTGACCCCG 1072
 941 GCTTACGCGGCTGCTGGGCTGCAACACAGACCGCTCAGCATATTCACGCGCTGT 1000
 1073 GATTGGCAAGGCTGCTGGAGTGACACACGAGCGAGCGGCTGATGACAGGCTGT 1132
 1001 GGCAGATGTGAAGACCAAGAGCTGAGAGTCCCATGCAAGATCATCATATGAGG 1060
 1133 GGGTTTACATCAGACCAACACAGCTGAGATGGGACAGCGGAGTACATTAATCGCA 1192
 1061 ACAAGTATTTCCAGAGATTTTGTGTCCTCCGCTGAGATTTTGTGAGATTCACG 1120
 1193 ACCGTACTTCGCGCAGATCTTCAATGTTGGCGACTCGGTTCTCGAGATTCACATGA 1252
 1121 GCTCAGAGCCCTGCTATGTCCTCCGAGCCCAATGTGATCAACATGTCATAGCTG 1180
 1253 AGCTGCGAGGTTGCTGAGCATCCAGACCCCAATGTGATCAACATGTCATAGCTG 1312
 1181 ACCCTCAGACGAGAGAGAGAGCGGTGATGATGATGCGTGGAGTGGAGAGCCAT 1240
 1313 ACCCTAAGACGAGAGAGAGAGCGCTGTTAGCATCGATGAGTGGAGTGGAGAGCCAC 1372
 1241 TAAAGGGGAGATGAGCAGCTTCTCTTATCCAGGCGCAACAGCAGAGATCACTCTC 1300
 1373 TGAAGGCCCAATGAGCAATTTCTGGCCCTTACCAACATGAGAGATGCGCTCCG 1432
 1301 TGGACAGTAAAGTCCATGAGAGATGAGTCCATTAACCACTCAAGATCCAGAGGACT 1360
 1433 TTGATGTCAAGATCCATGAGACCATGATGATCAACAGCAGTGAAGACCCCAAGAGATT 1492
 1361 TCATGTAGAGCTTCTCAGAGACCCCAAGGCTATGTCAGAGCCTGCTCCGCTCCACA 1420
 1493 TCATGCTAGTTTAAAGACGAGCCCGAGACTTCAATCAGAGATGAGCTCCGTTCCAG 1552
 1421 GCGGAGCTCAAGGATGATGAGATGATGAGCCGCAACCTGAAGAGAGCGCGGCTG 1480
 1553 GCGGAGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1612
 1481 AGTTTCAACCAAGCCCTGCTCCAGAGAGGCTCAGTGTGCTACTTCTACTGCAATCC 1540
 1613 CTTTTCAACCAAGCCCTGCTCCAGAGAGGCTCAGTGTGCTACTTCTACTGCAATCC 1672
 1541 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1597

Db 1673 ACCAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1729

RESULT 10

ABL02582

ABL02582 standard; cDNA; 4138 BP.

AC ABL02582;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2228.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila; gene; ss.

PN Drosophila melanogaster.

PD WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li P, Myers EW;

XX MPI, 2001-656860/75.

XX P-PSDB; ABB58479.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 2228; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins

XX CC (ABBS7737-ABBS2072).

XX The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4138 BP; 1065 A; 1001 C; 1024 G; 1048 T; 0 other;

XX Query Match 29.2%; Score 504.2; DB 23; Length 4138;

XX Best Local Similarity 66.6%; Pred. No. 2.3e-105; Indels 81; Gaps 4;

XX Matches 825; Conservative 0; Mismatches 333; Indels 81; Gaps 4;

433 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492

1649 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1708

493 TCCAGGCTTACATGAGCTCTTGGCATTTGAGAGGAACTGATCAACATCATGCGG 552

1709 TCGCAAGCTACATGATGATGCTGACGTTTGAAGCAAACTGATGCCAATATGCGC 1768

553 AAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 612

1769 AAGGCTTGGAGATCAGAGAGGCTCTCAAGCCCTCCATGAGAGCAAGAGGAGGAGGAGGAG 1828

613 CTCTATCTCAACACTTTTAACTGAGAGTGTGATGCTGAGAGATTCC---GACGAG 669

Dbb 1829 ATCTTTCATTCGAACACACATTCCTATCCCAAGAGAGCCCAACAAATGATGGCGAGGAGGC 1888
QY 670 AGCATTTGCTCTCTGGAGCTACGGGTGGAGGGAGAGCTCTGGATGATCCC-----720
Dbb 1889 GCCGTTGCTCTTTGGAAATTGGGGTGGAGGTGCCCTGTAGAGGATGGCAAGGGCCAT 1948
QY 721 -----AGCAAAACAAGCGGAAGTCTCTCTTTCTTCAAGAGTTTGGTTCATCGAGCTG 774
Dbb 1949 CCCAATACAAAGATCAACGCAAGTTTTCGTCGTTTTTCAAGTCTGTTATCGAGCTG 2008
QY 775 GACAAGATCTTTATGGCCCTGACAAACACCTCTGTAGTGCATCGACACCCACGACC 834
Dbb 2009 GATAGGAACCTGTACGGTCCGACAAACATCTGTGTCAGTGCATCGCACTCACACCAACC 2068
QY 835 CAGGAGCGGAGCGCTTCAGGTGAAACGGCTCGGGACCTGAGTGTGCGTGCACGCTG 894
Dbb 2069 CAGGAGCGGAGCGGTTTCAGGTGAAGCGCGCGCGATCGCAATGTGCGCTGCACCATC 2128
QY 895 CTCTTCATGCTGGACTACAGCCTCCCGAGTTCAAACCTGGATCCCGCTAGCCCGGCTG 954
Dbb 2129 CTCCTGCTGTGACTACAGCGCTGCAGTTCAAGCTGGACCGCGACTCGCAGGTTG 2188
QY 955 CTGGGGCTGCACACACAGAGCGCTCAGCCATTGTCCAGGCCCTGTGGCAGTATGTGAAG 1014
Dbb 2189 CTGGGCGTACACACACAGACAGCGCGGTGATCATATCCGCCCTGTGGCAGTACATCAAG 2248
QY 1015 ACCACAGCGCTGCAGACTCCCATCAAGGAATACATCAATGGGGACAAGTATTTCCAG 1074
Dbb 2249 ACGCAAGCTACAGGATGCCACAGCGGGAGTACATCAATTCGCAAGATATCTGGAG 2308
QY 1075 CAGATTTTGTGTTTCCCGGCTGAAGTTTCTGAGATTCCCGAGCGCTCACAGCCCTG 1134
Dbb 2309 CAGATATTTCAGTTGCGAGGATGAATTTGCCGAGATACCGCAACGCTCAATCCGCTG 2368
QY 1135 CTATTGCCCCCTGACCCCAATTGTCTATCAACATGTATCAGCGTGGACCCCTTCAGACAG 1194
Dbb 2369 CTGCATCCGCCCGATCCGATTTGTGATCAATCATTTTCAT---TGAGAGCGGTGCAGAGAAC 2425
QY 1195 AAGAGACAGCGTGTATGACATTACGCTGGAGGTGGAGGACCATTAAGGGGCGAGATG 1254
Dbb 2426 AAGCAGACTGCTCTGACATCATGTTGGAGGTGGATGATACGCTCAAAAACAGATG 2485
QY 1255 AGCAGCTTCTCTTATCCACGCGCCAAACAGCAGGAGATCAGTCTCTTGACAGTAAGATC 1314
Dbb 2486 AACAGCTTCTGATGAGCACTCGAGCCAAACAGGAGATCCAGGGCTGGACACCAAGATC 2545
QY 1315 CATGAGCAGTTAGTCCATAAACAGCTCAAGATCCAGAGGACTTCATGCTAAGCTTC 1374
Dbb 2546 CACGAGCGGTGGACACGATCAACACAGATGAAGACGAAACAGGGAGTTCTTCTAAGCTTC 2605
QY 1375 TCCAGAGACCCAAAGGCTATGTCCAAGACCTGCTCCGCTCCAGAGCGGGACCTCAAG 1434
Dbb 2606 GCCAAGGATCCAAATGTTTATTCATCGTGGATCATCAGTGAGACGAGGATTTGAAG 2665
QY 1435 G-----1435
Dbb 2666 GTAGTTCTTTTCTTTTTCCTTATACCGTTGCTAAACCTTTCGTCCTTCAG 2725
QY 1436 -TGATGACAGATGTAGCGGCAACCTGAAAGAGAGCGCGGCTGAGTTCTACACCAAG 1494
Dbb 2726 CTGATGACCGGATGTAGCTGGCAATCCGAGGAGGAGCGCTCGGGCGAGTTCTATTACCAG 2785
QY 1495 CCCTGGTCCAGGAGCGCTCAGTCTACTTCTACTCAAGATCCACAGCGGAGGAG 1554
Dbb 2786 CCATGAGCGCAGAGCGGCTTTCCGCGCTACTTCTTCCACCAAGTCAACAGAGCGGCGC 2845
QY 1555 GAGCTGGAGCAGTCTGCTGTTGTGGCGCAACACCTAGGAG 1593
Dbb 2846 GAATTTGAGAGCGCGCTGGCATAACGCAACGCTAGGTG 2884

AD53120 standard; DNA; 755 BP.
AAD53120;
28-MAY-2003 (first entry)
Human BRG1-associated factor (BAF) 60c3 DNA.
Human; nuclear receptor cofactor; proliferator-activated receptor; RXR;
PPAR-gamma; retinoid X receptor; cell proliferative disorder; infection;
metabolic disorder; cardiovascular disorder; inflammatory disease; acne;
neurodegenerative disease; Parkinson's disease; psoriasis; cancer; HIV;
renal disease; atherosclerosis; Alzheimer's disease; diabetes; vaccine;
osteoporosis; human immunodeficiency virus; BRG1-associated factor; BAF;
gene therapy; ds.
Homo sapiens.
Key Location/Qualifiers
CDS 1..753
FT /*tag= a
FT /product= "BAF60c3 protein"
FT /transl_except= (pos:598..609, aa:Arg-Gln)
FT /transl_except= (pos:610..618, aa:Met-Pro)
FT /note= "No stop codon"
FT /partial
XX WO200294877-A2.
XX PN 28-NOV-2002.
XX PD 23-MAY-2002; 2002WO-IB02939.
XX PF 23-MAY-2001; 2001US-292526P.
XX PR (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX PA Auwerx J, Gelman L, Debril M, Rocchi S, Picard F;
XX PI WPI; 2003-120789/11.
XX PS P-PSDB; AAE34721.
XX The invention relates to human nuclear receptor cofactor polypeptides.
XX The invention particularly relates to proliferator-activated receptor
XX (PPAR)-gamma and retinoid X receptor (RXR) cofactor polypeptides.
XX Polypeptides of the invention are useful for screening compounds that
XX modulate the interaction of the nuclear receptor with the nuclear
XX receptor cofactor. Modulators of the invention are useful in preparing
XX a pharmaceutical composition for treating and/or preventing diseases
XX or pathologic conditions associated with cell types that express PPAR
XX receptors. The pathologic conditions treated include metabolic or cell
XX proliferative disorders such as diabetes, cardiovascular disorders
XX (e.g. atherosclerosis), renal diseases, neurodegenerative diseases
XX (e.g. Parkinson's disease, Alzheimer's disease), inflammatory diseases
XX (e.g. psoriasis, acne), wounds, osteoporosis, infections (e.g. HIV)
XX or cancer. The invention is useful in gene therapy and as vaccines.
XX The present sequence is human BRG1-associated factor (BAF) 60c3 DNA.
XX This sequence is used in the exemplification of the invention.
SQ Sequence 755 BP; 159 A; 240 C; 247 G; 109 T; 0 other;
Query Match 27.0%; Score 466.2; DB 25; Length 755;
Best Local Similarity 98.8%; Pred. No. 7e-97; Mismatches 3; Gaps 3;
Matches 501; Conservative 0;


```

Db      78 GCGCCCGGAGATCGCTCTGAGAGCCGAGATGCCCAACAGAGGGGCGCCCATGGAGCCCCC 137
      278 GGGCTCCCGTATATAGGAGAGCCCGCCCTGCGAACCCGGGCTGGCCCCCGGGG-ATGGA 336
      138 GGGCTCCCGTATATAGGAGAGCCCGCCCTGCGAACCCGGGCTGGCCCCCGGGGCTGGA 197
      337 GCGCGCGCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 395
      198 GCGCGCGCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 257
      396 GCGGAGCGCCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455
      258 GCGGAGCGCCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 317
      456 AATCTCTCCCTCAAGAGATTGCGAGCTGGTCCCCGAGTCCGAGCTTACATGACCTCTT 515
      318 AATCTCTCCCTCAAGAGATTGCGAGCTGGTCCCCGAGTCCGAGCTTACATGACCTCTT 377
      516 GGCATTGAGAGGAACTGGATCAACATCATGCGGAGAGGGGGTGGACATCCAGAGAGG 575
      378 GGCATTGAGAGGAACTGGATCAACATCATGCGGAGAGGGGGTGGACATCCAGAGAGG 437
      576 TCTGAGAGGCGCCATGAGCAAAAGCGGAGGCTGCGACTCTATATCTCGAACACTTTTAA 635
      438 TCTGAGAGGCGCCATGAGCAAAAGCGGAGGCTGCGACTCTATATCTCGAACACTTTTAA 497
      636 CCTCTGGAAGTCTGATGCTGAGAGATTCCGAGCGGACGATTGCTCTGGAGGCTACGGGT 695
      498 CCTCTGGAAGTCTGATGCTGAGAGATTCCGAGCGGACGATTGCTCTGGAGGCTACGGGT 557
      696 GGAGGGGAGGCTCTCGATGATCCGAG 722
      558 GGAGGGGAGGCTCTCGATGATCCGAG 584

```

RESULT 12

ABT07117
ID ABT07117 standard; cDNA; 450 BP.

```

AC      ABT07117;
XX
DT      07-NOV-2002 (first entry)
XX
DE      Human ovarian cancer associated coding sequence SEQ ID NO: 279.
XX
KW      Human; ovarian cancer; cancer; gene; ss.
XX
OS      Homo sapiens.
XX
PN      US2002076715-A1.
XX
PD      20-JUN-2002.
XX
PF      06-JUN-2001; 2001US-0876889.
XX
PR      23-SEP-1998; 98US-0159320.
XX
PR      08-FEB-1999; 99US-0246429.
XX
PR      16-SEP-1999; 99US-0397787.
XX
PA      (BENS/) BENSON D R.
XX
PA      (LODE/) LODES M J.
XX
PA      (MITC/) MITCHAM J L.
XX
PA      (KING/) KING G E.
XX
PI      Benson DR, Lodes MJ, Mitcham JL, King GE;
XX
DR      MPI, 2002-598720/64.
XX
PT      Composition for detecting and treating ovarian cancer, comprises a
XX
PT      specific polypeptide, polynucleotide, T cell population, or antigen
XX
PT      presenting cell -
XX

```

PS Example 1; Page 105; 188bp; English.

XX The present invention relates to a method of detecting the presence of
CC ovarian cancer in a patient, involving detecting ovarian cancer
CC associated polynucleotides. The method is not only used to detect the
CC presence of cancer, preferably ovarian cancer in a patient, but also is
CC used to stimulate and/or expand T cells specific for an ovarian tumour
CC protein. The sequences can be used in vaccines used to treat cancer. The
CC present sequence is an ovarian cancer associated coding sequence.

Sequence 450 BP; 71 A; 170 C; 159 G; 48 T; 2 other;

Query Match 23.8%; Score 410.4; DB 24; Length 450;

Best Local Similarity 98.7%; Pred. No. 3.4e-84;

Matches 444; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

```

QY      40 GAGCGGCGCGCGCGCTCCGGGCGCGGGTCCCGGGGAGACAGATCTAGAAATGCCCCCTT 99
      1 GAGCGGCGCGCGCGCTCCGGGCGCGGGTCCCGGGGAGACAGATCTAGAAATGCCCCCTT 60
      100 GGTGCTGAGAGGCGCGTGGGCTCCGGGCGCGAGCGAGGGGGGAGCTGATGACTTCC 159
      61 GGTGCTGAGAGGCGCGTGGGCTCCGGGCGCGAGCGAGGGGGGAGCTGATGACTTCC 120
      160 AGTGCAGAGACCTTCCATCTATGACTCCAGTCTTTCAGACCCACCCAGCGTGTACAG 219
      121 AGTGCAGAGACCTTCCATCTATGACTCCAGTCTTTCAGACCCACCCAGCGTGTACAG 180
      220 CGCGCGGGGAGTCCGTGAGCGCGGATGCCCAACA-GGGGCGCCCATGGGCCCCCGG 278
      181 CGCGCGGGGAGTCCGTGAGCGCGGATGCCCAACAAGGGGGCGCCCATGGGCCCCCGG 240
      279 GCGTCCCGGTACATGGGAGAGCCCGCGGTGCGACCCCGGCTTGGCCCCCGGGG-ATGAG 337
      241 GCGTCCCGGTACATGGGAGAGCCCGCGGTGCGACCCCGGCTTGGCCCCCGGGGCTGGAG 300
      338 CGCGCGCGGAGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396
      301 CGCGCGCGGAGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
      397 CGGAGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 456
      361 CGGAGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
      457 ATCTCTCCCTCAAGAGATTGCGAGCTGGTC 486
      421 ATCTCTCCCTCAAGAGATTGCGAGCTGGTC 450

```

RESULT 13

ABX72995
ID ABX72995 standard; cDNA; 450 BP.

```

AC      ABX72995;
XX
DT      14-MAR-2003 (first entry)
XX
DE      Human ovarian carcinoma antigen partial cDNA sequence #261.
XX
KW      Human; cancer detection; ovarian carcinoma antigen; ovarian cancer;
XX
KW      tumour antigen; tumour; OV2; OV3; OV6; OV9; OV10; OV12; OV14; OV17;
XX
KW      OV18; OV23; OV24; OV27; OV41; OV54; OV57; gene; ss.
XX
OS      Homo sapiens.
XX
PN      US6468758-B1.
XX
PD      22-OCT-2002.
XX
PF      16-SEP-1999; 99US-0397787.
XX
PR      23-SEP-1998; 98US-0159320.
XX
PR      08-FEB-1999; 99US-0246429.
XX

```

XX (CORI-) CORIXA CORP.
XX Benson DR, Lodes MJ, Mitcham JL, King GE;
XX WPI; 2003-147101/14.
XX
XX Determining presence or absence of cancer in patient by contacting
XX patient sample with oligonucleotide that hybridises to polynucleotide
XX encoding ovarian carcinoma antigen, and detecting amount of
XX hybridisation in sample -
XX
XX Example 1; Column 201-202; 152pp; English.
XX
XX The present invention relates to compositions and methods for
XX determining presence or absence of cancer in patient. The method
XX comprising contacting a biological sample with an oligonucleotide
XX that hybridises to partial polynucleotide sequence encoding for
XX human ovarian carcinoma antigen, or its complement. The method and
XX compositions are useful for the therapy and diagnosis of ovarian
XX cancer. The method can be used to identify tumour antigens that
XX are secreted from ovarian carcinoma and/or other tumours. Effective
XX cancer detection is achieved using the method of the invention.
XX ABX72720-ABX73050 represent partial cDNA sequences encoding human
XX ovarian carcinoma antigens.
XX
XX Sequence 450 BP; 71 A; 170 C; 159 G; 48 T; 2 other;
XX
XX Query Match 23.8%; Score 410.4; DB 25; Length 450;
XX Best Local Similarity 98.7%; Pred. No. 3.4e-84;
XX Matches 444; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
XX
XX 40 GAGCGGGCGCGCGCTCCGGGCGGGGTCCGGGGAGCAGATCCTCAGAAATGSCCCTT 99
XX 1 GAGCGGGCGCGCGCTCCGGGCGGGGTCCGGGGAGCAGATCCTCANNATGGCCCTT 60
XX
XX 100 GGTGCTGAGGGCGCGGTGGGTCCGGGCGCCAGGCACCGAGGGGGCACTGGATCTCTCC 159
XX 61 GGTGCTGAGGGCGGTGGGTCCGGGCGCCAGGCACCGAGGGGGCACTGGATCTCTCC 120
XX
XX 160 AGGTGCAGGACCTGCCATCTATGACTCAGGTCTTACGACCCACCCACCGTGGTACAG 219
XX 121 AGGTGCAGGACCTGCCATCTATGACTCAGGTCTTACGACCCACCCACCGTGGTACAG 180
XX
XX 220 CGCCCGGGATCGCTCTGGAGCCCGGATGCCACCA-GGGGGCCCATGGCCCCCGG 278
XX 181 CGCCCGGGATCGCTCTGGAGCCCGGATGCCACCAAGGGGGCCCATGGCCCCCGG 240
XX
XX 279 GGCTCCCGGTACATGGGAGCGCCCGCGTGGACCCCGCTGGCCCGCCCGCGG-ATGGAG 337
XX 241 GGCTCCCGGTACATGGGAGCGCCCGCGTGGACCCCGCTGGCCCGCCCGCGGATGGAG 300
XX
XX 338 CCGCGCGGAGCGAGCGCGCCCGCCCGCC-GGCAGGCGGAGCAGAGCCAGGGCCAG 396
XX 301 CCGCGCGGAGCGAGCGCGCCCGCCCGCCCGGCGAGCGGAGCCAGAGCCAGGGCCAG 360
XX
XX 397 CGGAGGCGCCCGCCCGCGGAGCGCGAGTGCCTAAGAGGAGGAGAGTGGTGCACAAA 456
XX 361 CCGGTGCGCCCGCCCGCGGAGCGCGAGTGCCTAAGAGGAGGAGAGTGGTGCACAAA 420
XX
XX 457 ATCTCTCCCTCAAGAGATTTCGGAGCTGGTC 486
XX 421 ATCTCTCCCTCAAGAGATTTCGGAGCTGGTC 450
XX
XX RESULT 14
XX ID ABX47225
XX ABX47225 standard; cDNA; 285 BP.
XX
XX AC ABX47225;
XX
XX 21-FEB-2003 (first entry)
XX

DE Bovine EST associated with lactation/muscle/fat deposition #12390.
XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
XX muscle deposition; fat deposition; genome mapping; gene identification;
XX gene analysis; cattle breeding.
XX Bos Taurus.
XX OS
XX US2002137139-A1.
XX PN
XX 26-SEP-2002.
XX PD
XX 24-SEP-2001; 2001US-0960352.
XX PF
XX 12-JAN-1999; 99US-115707P.
XX PR
XX 11-JAN-2000; 2000US-0480902.
XX PA (BYAT// BYATT J C.
XX PA (MATH// MATHIALAGAN N.
XX PA (TAON// TAO N.
XX PA (WARR// WARREN W C.
XX
XX Byatt JC, Mathialagan N, Tao N, Warren WC;
XX WPI; 2003-110599/10.
XX DR
XX
XX New nucleic acid associated with lactation, and muscle and fat
XX deposition, useful for genome mapping, gene identification and
XX analysis, cattle breeding, or for genetically improving cattle -
XX
XX Claim 2; SEQ ID No 12390; 245pp; English.
XX
XX The invention relates to a purified nucleic acid molecule associated with
XX lactation or muscle and fat deposition (designated LMFD), derived
XX from cattle, and the LMFD nucleic acid can specifically hybridise to a
XX second nucleic acid molecule comprising any of 15112 nucleotide
XX sequences, appearing as ABX34836-ABX49947, or complements of them.
XX Also included are: (1) a transformed cell having a nucleic acid
XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-
XX translated sequence that functions in the cell to cause termination of
XX transcription and addition of polyadenylated ribonucleotides to a 3' end
XX of the mRNA molecule; and (2) determining a level or pattern of a
XX nucleic acid in a bovine cell or tissue comprising: (a) incubating a marker
XX nucleic acid (comprising any of the 15112 nucleic acid sequences or its
XX complement or fragment) with a complementary nucleic acid molecule
XX obtained from the bovine cell or tissue, where hybridisation between the
XX marker nucleic acid and the complementary nucleic acid permits the
XX detection of the molecule; and (b) detecting the level or pattern of the
XX complementary nucleic acid, where the detection of the complementary
XX nucleic acid is predictive of the level or pattern of the molecule.
XX The LMFD nucleic acid is used for determining a level or pattern
XX of a molecule in a bovine cell or tissue. It is useful for genome
XX mapping, gene identification and analysis, cattle breeding, preparation
XX of constructs for use in cattle gene expression, or for genetically
XX improving cattle. The present sequence is one of the 15112 bovine
XX LMFD EST (expressed sequence tag) nucleic acids.
XX Note: The present sequence was not shown in the specification but
XX was obtained in electronic format from the USPTO web site:
XX seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
XX Sequence 285 BP; 62 A; 79 C; 76 G; 68 T; 0 other;
XX
XX Query Match 12.0%; Score 206.2; DB 25; Length 285;
XX Best Local Similarity 83.0%; Pred. No. 1.4e-37;
XX Matches 235; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
XX
XX 559 GTGCATCCAGGAGGCTCTGAGAGGCCCATGAGCAAAAGCGGAAGCTCGCTCTAT 618
XX 3 GTGCTCATTCGTGATGCTCTGAAGCTGTTTACGAGCAGAGCGCAAGCTGGTGGCTAT 62
XX
XX 619 ATCTCCAACTTTTAACCCCTGCGAAGTCTGATCTGAGGATTCGAGCGGAGATTGCC 678
XX 63 ATCTCCAACTTTTAACCCCTGCTTAAGCCCGATGCTAGGATTCGAGCGGAGATTGCC 122
XX

0y	679	TCCTGGACCTACGGGTGGAGGGGAAAGCTTCCTGATGATCCGACAAACGAAGCCGAAG	738
Db	123	TCCTGGACCTCTGTGTGAGGGGAAAGCTTCCTGATGATCCGACCACTGCGCATATGCGGAAG	182
0y	739	TTCTCTCTTTCTTCAAGAGTTTGTCATCGACTGAGCAAAAGATCTTATGACCCTGAC	798
Db	183	TTCTCGTCTTCATCAAGAGCTCGGTGATTCGTGACATAGACCTTATGCCCCCTGAC	242
0y	799	AACCACTCTGTTGAGTGGCGATCGACACCCACGACCCAGAGA	841
Db	243	AACCACTGTTGAGTGGCGACCTGACACCCACCACTACAGAGA	285
RESULT 15			
ID	AAZ80712/C		
XX	AAZ80712	standard; CDNA; 452 BP.	
XX	AC		
XX	AAZ80712;		
DT	07-APR-2000	(first entry)	
DE	Human colon cancer cell line SW480	CDNA clone SEQ ID NO:796.	
XX			
KW	Human; gene expression product; diagnosis; tumour; colon cancer;		
KW	colorectal adenocarcinoma; cell line SW480; cell proliferation;		
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;		
XX	hyperplasia; ds.		
OS	Homo sapiens.		
XX			
PN	W09964576-A2.		
PD	16-DEC-1999.		
PE	09-JUN-1999;	99W0-IB01062.	
PR	10-JUN-1998;	98US-0088801.	
PA	(FARB) BAYER CORP.		
PI	Endege WO, Steimann KE, Aslie JH, Burgess CC, Bushnell SE;		
PI	Carroll E, Catino JV, Dertl A, Ford DM, Lewis ME, Monahan JE;		
PI	Schlegel R;		
DR	WPI; 2000-087220/07.		
PT	Novel nucleic acids, used to develop products for the diagnosis and		
PT	treatment of disorders involving unwanted cell proliferation,		
PT	particularly cancers, especially colon cancer -		
PS	Claim 15; Page 448; 469pp; English.		
XX			
CC	AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from		
CC	the human colorectal adenocarcinoma (colon cancer) cell line SW480. The		
CC	cDNA clones can be used to generate antisense oligonucleotides which		
CC	can be used for antisense therapy. Methods and products from the present		
CC	invention can be used for identifying and/or classifying cancerous cells		
CC	present in a human tumour, particularly in solid tumours, e.g.		
CC	carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones		
CC	can be used for developing agents for the diagnosis and treatment of		
CC	disorders involving unwanted cell proliferation, such as neoplasia,		
CC	dysplasia or hyperplasia.		
SO	Sequence 452 BP; 99 A; 124 C; 104 G; 122 T; 3 other;		

	Query Match	Similarity	Score	DB	Length
Best local	73.4%	Pred	No. 3	1e-37	452
Matches	312	Conservative	0	Mismatches	90
				Indels	23
				Gaps	3

	Query	Match	Similarity	Score	DB	Length
Oy	480	GCTGTCCTCCCGAGTCCAGGCTTACATGACCTCTTGCCATTTGAGAGGAACCTGATCA	539			
Ob	424	GCAGGTACACAGATCCAGGCTCAATATGATCTCTTGGCTTTTGAAGG-AACTGACCA	366			

QY	540	AACCAATTCGCGAAGGGGGGTGCACTCCAGAGGCTCTGAAGAGGCCCATGAAGCAAA	599
Db	355	GACTATCATGAGGAAC-GGCTAGATATCCAAAGGCTTTGAACGTCCTCATCAAGCAAAA	307
QY	600	GCGAAGCTCGACTCTATCTATCTCCACA	659
Db	306	ACGGAAGCTGGCAATTTTCATTTTCAACCTTTCAATCCGGCTAACTCAGATCCGAGGA	247
QY	660	TTCCGACGCGACGACTTGCCTCTGGGAGCTACGGGTGAGGGGAAAGCTCTCTG-	712
Db	246	TGGGGAAGGAGCGTGGCTCTCTGGAGCTTGGGGTAGAAGGACGGCTCTCGAGGATTC	187
QY	713	-----ATGATCCAGCAAAAGAGCGGAGTCTCTCTCTTTCTTCAAGAG	758
Db	186	AGCCTTTCCTCAATATGATGCACTAAACAAAGAGAAAGTTCTCTTTTAAATC	127
QY	759	TTTGGTCACTGAGCTGGAACAAGATCTTTATGAGCCCTGAACAACACCTGTTGAGTGCA	818
Db	126	CTTGTGATTGAACGTGGAACAAGACTGTGATGGGCGAGAACCACTGCTGTAGAAATGGCA	67
QY	819	TGGAACACCCACGACCCAGAGAAAGGAGCGGCTTCCAGGTGAACCGGCTTGGGGACCTGAG	878
Db	66	CAGGACCGCACTACCCAGAGACCGATGGCTTTAGGTGAAGCGGCGCGGAGACGCTGAA	7
QY	879	TGTGC	883
Db	6	TGTAC	2

Search completed: October 11, 2003, 09:23:56
Job time : 367 secs

THIS PAGE BLANK (USPTO)

Qy 338 CCGGCCCGCAAGCAGCAGCGCCCGCCCGCC-GGCAGAGCCAGGCAAGAGCCAGGCGCAG 396
Db 301 CCGGCCCGCAGCAGCAGCGCCCGCCCGCCCGGCGCAGGCGCAGGCGCAG 360
Qy 397 CCGGAGCCACCGCCCGCCCGCGGAGCCGCGAGTCGCAAGAGAGGAGAAATGGCTGACAAA 456
Db 361 CCGGTGCCACCGCCCGCCCGCGGAGCCGCGAGTCGCAAGAGAGGAGAAATGGCTGACAAA 420
Qy 457 ATCTCTCCCTCAAGGATTGCGSAGCTGGTC 486
Db 421 ATCTCTCCCTCAAGGATTGCGGAGCTGGTC 450

RESULT 2

US-09-328-111-796/c
; Sequence 796, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 796
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-796

Query Match 11.9%; Score 205; DB 3; Length 452;
Best Local Similarity 73.4%; Pred. No. 1.3e-40;
Matches 312; Conservative 0; Mismatches 90; Indels 23; Gaps 3;
Qy 480 GCTGTCCTCCGAGTCCCGAGCTTACATGAGACCTCTTGGCATTTGAGAGAACTGGATCA 539
Db 424 GCAGGTACCAGAAATCCCGAGGCTATATGATCTCTTGGCTTTTGAAGG-AACTGGACCA 366
Qy 540 AACCATCATGCGAAGGGGTGGACATCCAGGAGGCTCTGAGAGGCCCATGAAGCAAAA 599
Db 365 GACTATCATGAGGAAC-GGCTAGATATCAAGAGGCGCTTGAACGCTCCCATCAAGCAAAA 307
Qy 600 GGGGAAGTGGGACTCTATATCTCCACACTTTTAACTCTGCAAGTCTGATGCTGAGGA 659
Db 306 ACGGAAGTGGGAATTTTCATTTCTTAACTTCAATCGGTAAGTCAGATGCGGAGGA 247
Qy 660 TTCCGACCGCAGCATTTGCTCTCTGGAGCTACGGGTGAGGGGAAAGCTCTCTGG----- 712
Db 246 TGGGAAGGAGCGGTGGCTTCTGGAGCTTCGGGTAGAAAGGAGCGGCTCTGGAGGATTC 187
Qy 713 -----ATGATCCAGCAAAACAGNAGCGGAAGTTCTCTTCTTCTTCAAGAG 758
Db 186 AGCCTTGTCAAATAATGATGCCATTAACAAAGAGGAAGTTCTCTCTCTTTTAAATC 127

Qy 759 TTTGGTCAATCAGCTGCACAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTGGCA 818
Db 126 CTTGGTGATTGAACCTGGCAAAAGACCTGTATGGGCCAGACAAACCATCTGGTAGAATGGCA 67
Qy 819 TCGGACACCCACGACCCAGGAGCGGCTTCCAGGTGAAACGGGCTGGGACCTGAG 878
Db 66 CAGGACCGCCACTACCCAGGAGACCGATGGCTTTTCAGGTGAAGCGGCGGAGACGTGAA 7
Qy 879 TGTGC 883
Db 6 TGTAC 2

RESULT 3

US-09-328-111-645/c
; Sequence 645, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 645
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(690)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-645

Query Match 10.4%; Score 179.2; DB 3; Length 690;
Best Local Similarity 64.5%; Pred. No. 2.7e-34;
Matches 323; Conservative 0; Mismatches 169; Indels 9; Gaps 4;
Qy 1112 TTCCCCAGCGCCTCACAGCCCTGCTATTGGCCCCGTGACCCCAATTGTCAATCA-ACCATGTC 1170
Db 638 TCCCTAAGNGGTTCCCATNCNTGTTAATGCGCNCANAACTATTNTCTTATTCAATGTC 579
Qy 1171 ATCAGCTGGACCCCTTCAGACCCAGAGAA-GACAGCGTGTCTATCATTGACGTGGAGGT 1229
Db 578 ATCAGTGTGACCCGAATGTTNAGAAAAAGCCAGCTTGTATTGCCATTGATGTTNAAAG 519
Qy 1230 GGAGGAGCCATTAAAGGGGAGATGAGCAGCTTCTCT-----CCTATCCAGCGGCAACCA 1283
Db 518 TGATGCGCCTTGAAGAGCCCAAGATGATTCTTTCTGCTGCTCCANTGCCGGGCCA 459
Qy 1284 GCAGGAGTCACTCTCTGGACAGTAAGATCCATGAGACGATGAGT-CCATTAACCCAGC 1342
Db 458 CCAGAAGATTGGTACTTTAGCCACCAAGATCCATGAGCAATAGAAAACCCCATCAACCCAGC 399
Qy 1343 TCAAGATCCAGAGGACTTCATGCTAAGCTTCTCCAGAGACCCCAAGAGGCTATCTCCAG 1402
Db 398 TGAAGACTCAGCGGGAGTTTATGTTAAGTTTGGCCAGAGACCCCTCAGGGTTTCATCAATG 339


```
QY 228 GATGCGCTGTGAGCCGCGATGCCCCACAGGGGGCCCATGGGCGCCCGGCTTCCCG 287
    |||||
Db 927077 TCCGCGCGCCCGCGGCCCGGAAACAGCCACCGACCGCGCCCGCGCCCGCC 927018
    |||||
QY 288 TACATGGGAGAGCCCGCGCTGTGACCGCGCCTTGGCCCGCGGGATGAGCCCGCCGCA 347
    |||||
Db 927017 GACGTGTTGTCGCGATCCCGCGCCCGCGCGCCCGCGCCCGCGCGCAACAGCCACCGC 926958
    |||||
QY 348 ACGGAGCAGCGCCCGCGCCCGCGCAGAGCCAGGAGCCAGCGCGAGCCCGCCAC 407
    |||||
Db 926957 CCGCGCGCGCGCGCCCTTCCCGCGCAGCCCGCGCTGACGCGGACCGCGCGCCCGCC 926898
    |||||
QY 408 CGCCCCCGCGCGAGCCGCGAG 428
    |||||
Db 926897 GCGCCCGCGCGCGCTTGAACAG 926877
    |||||
```

```
RESULT 7
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16
```

```
Query Match 2.7%; Score 46.4; DB 3; Length 152331;
Best Local Similarity 50.0%; Pred. No. 0.31; Mismatches 116; Indels 0; Gaps 0;
Matches 116; Conservative 0;

QY 201 CCCACCCACCGTGTACAGCGCCCGGGATGCGCTGTGAGCGCCGGATGCCCCACAGGG 260
    |||||
Db 21964 CCCCCACCCCCCCCCCCCCCCCCCGCGCGCCCGCCCCCGCGCCCGCCCCACCCCC 22023
    |||||
QY 261 GCGCCCATGGGCCCCCGGGCTTCCCGTACATGGGAGCGCCCGCGTGTGACCGCGCTG 320
    |||||
Db 22024 CCGCCCCCGCGCCCCCCCCCCCCCCCCCCCCACACCCCGCGCCACAGCACCCCCCCC 22083
    |||||
QY 321 GCGCCCGGGGTAGAGCCCGCGCGCGCGAGCAGCGCCCGCGCGCGCGCGCGCGCGCGC 380
    |||||
Db 22084 ACCCGGAGCGCCCGCCCCCCCCCCCCCCCCCGCGCAGCGCGCCCGCCCCCGCGCCCG 22143
    |||||
QY 381 ACAGAGCAGGCGCCAGCGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
    |||||
Db 22144 CCGCACCCCGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 22195
    |||||
```

```
RESULT 8
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
```

```
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 2.7%; Score 45.8; DB 3; Length 4403765;
Best Local Similarity 50.2%; Pred. No. 1.5;
Matches 113; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 196 CAGCACCACCCACCGTGTACAGCGCCCGGGATGCCCGTCTGGAGCCCGGATGCCCCAC 255
    |||||
Db 675058 CCGCGCGACCGCGCGGCGCTTGGCTGCGCGGACGCGCGTGGGTGCGCGCCCGCGCTCC 675117
    |||||
QY 256 CAGGGCGCCCATGGCCCCCGGGCTCCCGTACATGGGAGCGCCCGCGCGTGCACCCG 315
    |||||
Db 675118 CCGCGTCCCCACCTTTTCCGCGGTACCGCAACTCCGCGGTGCGCGCGGGGTGCCG 675177
    |||||
QY 316 GCCTGGCCCCCGGGATGGAGCCCGCGCAAGCAGCAGCGCGCCCGCGCGCGCGCGAGC 375
    |||||
Db 675178 TCCGCGCGCGCTGGAACCGTTGACACCGTGGTGGCGGACCTCCAGTCCCGCGGAG 675237
    |||||
QY 376 CAGGCACAGAGCCAGCGCGAGCCCGAGCCCAACCGCCCCCGCGCGCG 420
    |||||
Db 675238 CCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675282
    |||||
```

```
RESULT 9
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match 2.7%; Score 45.8; DB 3; Length 4411529;
Best Local Similarity 50.2%; Pred. No. 1.5;
Matches 113; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 196 CAGCACCACCCACCGTGTACAGCGCCCGGGATGCCCGTCTGGAGCCCGGATGCCCCAC 255
    |||||
Db 673615 CCGCGCGACCGCGCGGCGCTTGGCTGCGCGGACGCGCGTGGGTGCGCGCCCGCGCTCC 673674
    |||||
QY 256 CAGGGCGCCCATGGCCCCCGGGCTCCCGTACATGGGAGCGCCCGCGCGCGCGTGCACCCG 315
    |||||
```


Db 673735 CCGCGGTCCTCCACCTTTTCCCGGTCACCCCACTCCCGCGGCGCGGCGGCGCG 673734
Qy 316 GCGTGGCCCCCGGGGATGAGACCCCGCGCAGAGAGAGCGCCCGCGCGGAGAGC 375
Db 673735 TCCGGCGCCCGCTGACCGTTGACACCGTCTGCGGAGCCCTCCAGTCCCGCGAGC 673794
Qy 376 CAGCAGACAGAGCCAGCGAGCGGAGCCGAGCCGCGCGCGCGCGCG 420
Db 673795 CCGCGGTCCTCCCGCGGCGGTCGACCGCTTCCCGCGCGAG 673839

RESULT 10
US-07-928-611-12
Sequence 12, Application US/07928611
Patent No. 5569601
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: A No. 5569601e1 Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/928,611
FILING DATE: 19920810
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5569601nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 1..803
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "Alternate Exon 3: D4.7"
OTHER INFORMATION: /note= "This sequence represents the third exon of
OTHER INFORMATION: allele D4.7 of the human D4 dopamine receptor
OTHER INFORMATION: gene"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 257..262
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "PstI site"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "PstI site"
OTHER INFORMATION: /label= PstI
OTHER INFORMATION: /note= "This sequence is a PstI site whereby
OTHER INFORMATION: digestion of human genomic DNA produces a RFLP"
FEATURE:
NAME/KEY: repeat_region
LOCATION: 346..682

IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /rpt_type= "tandem"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /rpt_unit= 346..394
OTHER INFORMATION: /note= "This sequence is a repeat found in 7 known
OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
FEATURE:
NAME/KEY: CDS
LOCATION: 2..803
US-07-928-611-12

Query Match 2.6%; Score 44.8; DB 1; Length 803;
Best Local Similarity 47.2%; Pred. No. 0.11;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2;

Qy 7 CCGCGCAGCGCCCGGAGCGGCGGAGCGAGGAGCGGCGCGCGCTCCGCGCGG 66
Db 281 CGCGCAGCTGACAGCGCGCGCGCGCGCGCGCGCGAGCCGCGCTTCCCGCG 340
Qy 67 GTCCGGGGGAGCAGATGCTCAGATGCGCTTGCTGTGAGAGCGCGGTGCGCGG 126
Db 341 AGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCTGTGCGCGCG 400
Qy 127 CCGAGCAGCAGGAGGCGCACTGATGATCTTCAGAGTCAGAGACCTGCACTATGACT 186
Db 401 CCGGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 459
Qy 187 CGAGTCTTACAGACCCAGCCAGCCGCTGTAGAGCGCGCGCGGATGCGCTTGGAGCCCGG 246
Db 460 CCGGAGCCCTGCG 519
Qy 247 ATGCCCCAGAGGCG 306
Db 520 CGCGCGCGCTGTGCTG 579
Qy 307 TCGGACCC--GGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 363
Db 580 TCG 639
Qy 364 GCGCGCAGAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 423
Db 640 GCGCGCGCTCCCG 699
Qy 424 CGCAGTGC 431
Db 700 CGCGCGCG 707

RESULT 11
US-08-487-811A-12
Sequence 12, Application US/08487811A
Patent No. 5883226
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: A No. 5883226e1 Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,811A
FILING DATE: 07-JUN-1995


```

FEATURE:
NAME/KEY: repeat region
LOCATION: 346..682
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /ipc_type="tandem"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /ipc_unit=346..394
OTHER INFORMATION: /note="This sequence is a repeat found in 7 known
OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
FEATURE:
NAME/KEY: CDS
LOCATION: 2..803
US-09-060-694-12

Query Match
Best Local Similarity 47.2%; Score 44.8; DB 3; Length 803;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2;

QY 7 CGGCGGAGGCGCGCGAGCGCGGAGCGGAGCGGCGGCGCGCTCCGCGCGG 66
DB 281 CGCGCTAAGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCC 340
QY 67 GTCCGCGGAGCGCGAGATCTCTGAGATGCGCTTGTGCTGACAGCGCGGCTCGG 126
DB 341 AGGCGACCGCGCGCGCGCGCTCCCGCGAGACCGCTGCGCGCGCGAGCTGCGCGCG 400
QY 127 CCAGGACCGAGGCGGCGCTGAGTACTCTCCAGTGCAGAGACCTGCGCATATGACT 186
DB 401 CCGGCGCTTCCCGGCGGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 459
QY 187 CGAGCTTCTCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
DB 460 CCGGACCGCTGCGCGCGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 519
QY 247 ATGCCCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
DB 520 CGGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579
QY 307 TCGAGCC--GCGTGTGCGCGCGCGCGAGTGAAGCGCGCGCGCGAGCGCGCG 363
DB 580 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
QY 364 GCGCGCGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423
DB 640 GCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
QY 424 CGCAGTGC 431
DB 700 CGCGCGCG 707

RESULT 13
US-09-378-074-12
Sequence 12, Application US/09378074
Patent No. 6437114
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
Civelli, Olivier
TITLE OF INVENTION: A No. 6437114el Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegrucci & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/378,074
FILING DATE: 20-AUG-1999
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/928,611
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No. 6437114nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 1..803
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
/standard_name="Alternate Exon 3: D4.7"
/note="This sequence represents the third exon of
allele D4.7 of the human D4 dopamine receptor
gene"
FEATURE:
NAME/KEY: misc feature
LOCATION: 257..262
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="PstI site"
/evidence=EXPERIMENTAL
/standard_name="PstI site"
/label=PstI
/note="This sequence is a PstI site whereby
digestion of human genomic DNA produces a RFLP"
FEATURE:
NAME/KEY: repeat region
LOCATION: 346..682
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /ipc_type="tandem"
/evidence=EXPERIMENTAL
/ipc_unit=346..394
/note="This sequence is a repeat found in 7 known
alleles of the human D4 dopamine receptor gene
encoding a 16 amino acid sequence repeated 7 times
FEATURE:
NAME/KEY: CDS
LOCATION: 2..803
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-378-074-12

Query Match
Best Local Similarity 47.2%; Score 44.8; DB 4; Length 803;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2;

QY 7 CGGCGGAGGCGCGCGAGCGCGGAGCGGAGCGGCGGCGCGCTCCGCGCGG 66
DB 281 CGCGCTAAGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCC 340
QY 67 GTCCGCGGAGCGCGAGATCTCTGAGATGCGCTTGTGCTGACAGCGCGGCTCGG 126
DB 341 AGGCGACCGCGCGCGCGCGCTCCCGCGAGACCGCTGCGCGCGCGAGCTGCGCGCG 400
QY 127 CCAGGACCGAGGCGGCGCTGAGTACTCTCCAGTGCAGAGACCTGCGCATATGACT 186
DB 401 CCGGCGCTTCCCGGCGGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 459

```

```

187 CCAGGTCTTACGACACCCACCCAGCTGGTACAGCGCCCCCGGATGCCGTCTGGAGCCCGG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 CCCGACCCCTGGGCCCCGAGTGTGCGCCCCCGCGCCCTCCCGCCAGGACCCCTG 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 ATGCCCAACAGGGGCGCCATGGGCCCCCGGGTCCCGGTACATGGGCGAGCCCCGCG 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 CGGCCCGGACTGTGCGCCCCCGCGCCCGGCTTCCCGGGGTCCCTCGGCGCCGACTG 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
307 TCGGACCC---GGCTGGCCCCCGGGATGGAGCCCCCGCGAAGCGAGCGCCCC 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 TCGGCCCCCGCGCCCGGCTCCCCAGAGACCCCTTCGCGGCCGCGACTGTGCGCCCCCGC 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 GCCCGGAGAGCGAGCGACAGAGCCAGCGGCGAGCGCCAGCCCGCCCCCGCGGAGC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
640 GCCCGGCTCCCGCGGACCCCTGCGGCTCCAACGTGCTCCCCCGGACGCGTCAGAGC 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 CGCAGTGC 431
    ||| ||| |||
700 CGCGCGC 707
    ||| ||| |||

RESULT 14
PCT-US93-07370-12
; Sequence 12, Application PC/TUS9307370
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A Novel Human Dopamine Receptor and Uses
; NUMBER OF SEQUENCES: 22
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: PCT/US93/07370
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1..803
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "Alternate Exon 3: D4.7"
; OTHER INFORMATION: /note= "This sequence represents the third exon of
; OTHER INFORMATION: allele D4.7 of the human D4 dopamine receptor
; OTHER INFORMATION: gene"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 257..262
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "PstI site"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "PstI site"
; OTHER INFORMATION: /label= PstI
; OTHER INFORMATION: /note= "This sequence is a PstI site whereby
; OTHER INFORMATION: digestion of human genomic DNA produces a RFLP"
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 346..682
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /rpt_type= "tandem"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /rpt_unit= 346..394
; OTHER INFORMATION: /note= "This sequence is a repeat found in 7 known
; OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
; OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: 2..803
PCT-US93-07370-12
Query Match 2.6%; Score 44.8; DB 5; Length 803;
Best Local Similarity 47.2%; Pred. No. 0.11; 22; Indels 4; Gaps 2;
Matches 202; Conservative 0; Mismatches 0;

QY 7 CGGCGCAGGGCGCCGAGCCGAGCCGAGCAGGAGGCGGCGCGCTCCGGGCGCGG 66
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 CGGCCCAAGCTGACGGCGCGCGCCCGCGAGCCAGCGCCCTGGCCCCGCTTCCCC 340
QY 67 GTCCCGGGGAGCAGATCCTCAGAAATGGCCCTTGGTGTGAGGGCGGCGTGGGTCCGGG 126
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 ACGCCACCCGCGCCCGCTCCCGCAGGACCCCTGCGGCGCCGAGCTGTGCGCCCCCGCG 400
QY 127 CCAGGACACGAGGGGCACTGGATGACTCTCCAGGTGCGAGGACCTGCCATCTATGACT 186
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 CCGGCGCTTCCCGGGGTCCCTGCGGCCCCGAC-TGTGCGCCCCCGCGCCCGGCTCCC 459
QY 187 CCAGGTCTTACGACACCCACCCAGCTGTAGCGCCCCCGGGATGCCGTCTGGAGCCCGG 246
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 CCGGACCCCTGCGGCCCCGACTGTGCGCCCCCGCGCCCTCCCGCAGGACCCCTG 519
QY 247 ATGCCCAACAGGGGCGCCATGGGCCCCCGGGCTCCCGGTACATGGGCGAGCCCCGCG 306
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 CGGCCCGGACTGTGCGCCCCCGCGCCCGGCTTCCCGGGGTCCCTCGGCGCCGACTG 579
QY 307 TCGGACCC---GGCTGGCCCCCGGGATGGAGCCCCCGCGAAGCGAGCGCCCC 363
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 TCGGCCCCCGCGCCCGGCTCCCCAGGACCCCTTCGCGGCCGCGACTGTGCGCCCCCGC 639
QY 364 GCCCGGAGAGCGAGCGACAGAGCCAGCGGCGAGCGCCAGCCCGCCCCCGCGGAGC 423
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
640 GCCCGGCTCCCGCGGACCCCTGCGGCTCCAACGTGCTCCCCCGGACGCGTCAGAGC 699
QY 424 CGCAGTGC 431
Db ||| ||| |||
700 CGCGCGC 707
    ||| ||| |||

RESULT 15
US-09-016-434-1276
; Sequence 1276, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071

```

REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1276:
SEQUENCE CHARACTERISTICS:
LENGTH: 1504 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 9291945
US-09-016-434-1276

Query Match 2.6%; Score 44.8; DB 4; Length 1504;
Best Local Similarity 47.2%; Pred. No. 0.14;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2;
QY 7 CGAGCGAGGCGCCGAGCCGAGCGGAGCGAGGAGGAGCGGCGCGCTCCGGGCGGG 66
DB 679 CGGCGCAAGCTGACG 738
QY 67 GTCCCGGGGAGCAGATCTCAAAATGACCTTGTGTGTCAGAGCGGCTGGGCTCCGGG 126
DB 739 AGGCGACCG 798
QY 127 CCCAGGACCGAGGGGAGCACTGATGACTCTCCAGGTGAGAGCCTGCCATCTATGACT 186
DB 799 CCGGCGCTTCCCGGGGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 857
QY 187 CCAGGCTTCAAGACCCACCAACCGTGTACAGCGCGCGCGGATGCGTGTGAGCGCGG 246
DB 858 CCGGACCCCTGCGGCG 917
QY 247 ATGCCCCACGAGGGGCGCCATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
DB 918 CGGCGCGGACTGAGCG 977
QY 307 TCGGAGCC--GGCTGGCGCGCGCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 363
DB 978 TCG 1037
QY 364 GCCCGGCGAGGCGGACAGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 423
DB 1038 GCGCGGCTTCCCG 1097
QY 424 CGCAGTGC 431
DB 1098 CGCGCGCGC 1105

Search completed: October 11, 2003, 11:20:04
Job time : 117 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 10:35:17 ; Search time 351 Seconds
(without alignments)
12744.885 Million cell updates/sec

Title: U66619

Perfect score: 1724
Sequence: 1 GAATTCGCGCGAGCGGCC.....GTCATTATCTGGAATTC 1724

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	450.8	26.1	485	11	US-09-918-995-16164
2	421.2	24.4	462	11	US-09-918-995-28488
3	412.2	23.9	474	11	US-09-918-995-26465
4	410.4	23.8	450	9	US-09-876-889-279
5	346	20.1	447	11	US-09-918-995-28419
6	340.4	19.7	471	11	US-09-918-995-29605
7	206.2	12.0	285	10	US-09-960-352-12390
8	205	11.9	452	10	US-09-879-536-796
9	203	11.8	495	11	US-09-918-995-2832
10	180.4	10.5	420	11	US-09-918-995-34033
11	179.2	10.4	690	10	US-09-879-536-645
12	170.4	9.9	8788	10	US-09-764-877-2295
13	170.4	9.9	8788	10	US-09-860-670-247
14	162	9.4	440	11	US-09-918-995-27234
15	137.8	8.0	310	14	US-10-070-676-23
16	132	7.7	520	9	US-09-864-761-8048

C	17	129.2	7.5	219	9	US-09-864-761-24793	Sequence 24793, A
	18	91.6	5.3	15500	9	US-09-764-869-2109	Sequence 2109, Ap
	19	91.6	5.3	15500	14	US-10-091-504-2109	Sequence 2109, Ap
	20	84.6	4.9	2017	14	US-10-102-806-137	Sequence 137, App
	21	82.8	4.8	508	11	US-09-918-995-997	Sequence 997, App
	22	67.8	3.9	215	10	US-09-867-701-9486	Sequence 9486, Ap
	23	66.2	3.8	740	9	US-09-770-149-60	Sequence 60, Appl
	24	60	3.5	60	12	US-09-908-975-6179	Sequence 6179, Ap
	25	53	3.1	9025608	14	US-10-156-761-1	Sequence 1, Appli
	26	50.6	2.9	1437	14	US-10-128-714-2188	Sequence 2188, Ap
	27	50.6	2.9	1437	14	US-10-128-714-1188	Sequence 7188, Ap
	28	50.6	2.9	1493	14	US-10-128-714-1188	Sequence 1188, Ap
	29	50.6	2.9	1493	14	US-10-128-714-5188	Sequence 6188, Ap
	30	50.6	2.9	3340	14	US-10-128-714-187	Sequence 187, App
	31	50.6	2.9	3340	14	US-10-128-714-5187	Sequence 5187, App
	32	50.6	2.9	3493	14	US-10-128-714-188	Sequence 188, App
	33	50.6	2.9	3493	14	US-10-128-714-5188	Sequence 5188, Ap
	34	50.2	2.9	4908	13	US-10-001-887-33	Sequence 33, Appl
	35	49.8	2.9	28000	12	US-10-091-625-11	Sequence 11, Appl
C	36	49.8	2.9	28000	12	US-10-096-399A-11	Sequence 11, Appl
	37	49.4	2.9	522	13	US-10-027-632-129532	Sequence 129532,
	38	48.6	2.8	2112	14	US-10-219-449-3	Sequence 3, Appli
	39	48.6	2.8	2154	14	US-10-219-449-1	Sequence 1, Appli
	40	47.2	2.7	2678	12	US-10-017-161-1647	Sequence 1647, Ap
	41	47	2.7	2561	10	US-09-976-740-48	Sequence 48, Appl
	42	47	2.7	2561	13	US-10-023-529-48	Sequence 48, Appl
	43	47	2.7	2561	13	US-10-023-523-48	Sequence 48, Appl
	44	46.6	2.7	154746	12	US-09-827-688-8	Sequence 8, Appli
C	45	46.6	2.7	154746	12	US-09-827-688-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-09-918-995-16164
; Sequence 16164, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/918,995
; PRIORITY FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16164
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(485)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-16164

Query Match
Best Local Similarity 98.5%; Pred. No. 5.6e-120;
Matches 455; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 710 TGGATGATCCAGCAAGAGGAGTTCTTTCTTTCTTCAAGATTGTCATCG 769
23 TGGAGCACTCAGCAGACAGAAAGCGAGTTCTTTCTTTCTTCAAGATTGTCATCG 82
QY 770 AGCTGACAAAGATTTTATGCGCTTGACACACACTCTGTGAGTGGCATCGAACCACCA 829
83 AGCTGACAAAGATTTTATGCGCTTGACACACACTCTGTGAGTGGCATCGAACCACCA 142
QY 830 CGACCAGAGACGAGCGGCTTCCAGTGAACGCGCTGGGACCTGAGTGGCCTGCA 889
```


RESULT 4
 US-09-876-889-279
 ; Sequence 279, Application US/09876889
 ; Patent No. US20020076715A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benson, Darin R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: King, Gordon E.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
 ; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
 ; FILE REFERENCE: 210121.466C3
 ; CURRENT APPLICATION NUMBER: US/09/876,889
 ; CURRENT FILING DATE: 2001-06-06
 ; NUMBER OF SEQ ID NOS: 353
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 279
 ; LENGTH: 450
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(450)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-876-889-279

Query Match 23.8%; Score 410.4; DB 9; Length 450;
 Best Local Similarity 98.7%; Pred. No. 2,6e-108;
 Matches 444; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 40 GAGCGGCGCGCGCGCTCCGCGGCGCGGAGATCCGAGATGCGCCCTT 99
 DB 1 GAGCGGCGCGCGCGCTCCGCGGCGCGGAGATCCGAGATGCGCCCTT 60
 QY 100 GGTGCTGACAGCGCGGTGGGCTCCGCGGCGCGGAGATCCGAGATGCGCCCTT 159
 DB 61 GGTGCTGACAGCGCGGTGGGCTCCGCGGCGCGGAGATCCGAGATGCGCCCTT 120
 QY 160 AGGTGACAGACCTGCTCATATGATCCAGGTTCTTACAGACCCACCGCTGTAACAG 219
 DB 121 AGGTGACAGACCTGCTCATATGATCCAGGTTCTTACAGACCCACCGCTGTAACAG 180
 QY 220 CGCGCGGAGATGCGGTGGAGCGCGGATGCCACCA-GGGCGCGCATGGGCGCGCGG 278
 DB 181 CGCGCGGAGATGCGGTGGAGCGCGGATGCCACCA-GGGCGCGCATGGGCGCGCGG 240
 QY 279 GGTGCTGACAGACCTGCTCATATGATCCAGGTTCTTACAGACCCACCGCTGTAACAG 337
 DB 241 GGTGCTGACAGACCTGCTCATATGATCCAGGTTCTTACAGACCCACCGCTGTAACAG 300
 QY 338 CGCGCGGAGATGCGGTGGAGCGCGGATGCCACCA-GGGCGCGCATGGGCGCGCGG 396
 DB 301 CGCGCGGAGATGCGGTGGAGCGCGGATGCCACCA-GGGCGCGCATGGGCGCGCGG 360
 QY 397 CGCGAGCG 456
 DB 361 CGCGAGCG 420
 QY 457 ATCTCCCTCAAGAGATTGGAGCTGTGC 486
 DB 421 ATCTCCCTCAAGAGATTGGAGCTGTGC 450

RESULT 5
 US-09-918-995-28419
 ; Sequence 28419, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 28419
 ; LENGTH: 447
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(447)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-28419

Query Match 20.1%; Score 346; DB 11; Length 447;
 Best Local Similarity 94.7%; Pred. No. 1e-89;
 Matches 375; Conservative 0; Mismatches 5; Indels 16; Gaps 1;

QY 498 GGCTTACATGACCTCTTGGCATTTGAGAGAACTGGATCAACCATGCGGAGGG 557
 DB 52 GGAATGATGACCTCTTGGCATTTGAGAGAACTGGATCAACCATGCGGAGGG 111
 QY 558 GGTGACATCCAGAGGCTCTGAAGAGGCCATGAAGCAAAAGCGAAGCTGCACTTA 617
 DB 112 GGTGACATCCAGAGGCTCTGAAGAGGCCATGAAGCAAAAGCGAAGCTGCACTTA 171
 QY 618 TATCTTCAACACTTTTAACCTTGGCGAAGTCTGATGCTGAGATTCCGACGAGACTTC 677
 DB 172 TATCTTCAACACTTTTAACCTTGGCGAAGTCTGATGCTGAGATTCCGACGAGACTTC 231
 QY 678 CTCTTGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 721
 DB 232 CTCTTGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 291
 QY 722 GCAACAGAGCGAAGTCTCTTCTTCTTCAAGATTGCTCATGAGCTGACAAAG 781
 DB 292 GCAACAGAGCGAAGTCTCTTCTTCTTCAAGATTGCTCATGAGCTGACAAAG 351
 QY 782 ATCTTATGAGCCTGACCAACCACTGTTGATGATGATGATGATGATGATGATGATGAT 841
 DB 352 ATCTTATGAGCCTGACCAACCACTGTTGATGATGATGATGATGATGATGATGATGAT 411
 QY 842 CGAGCGGCTTCCAGGTGAACGCGCTGGGAGACTGA 877
 DB 412 CGAGCGGCTTCCAGGTGAACGCGCTGGGAGACTGA 447

RESULT 6
 US-09-918-995-29605
 ; Sequence 29605, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 29605
 ; LENGTH: 471
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(471)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-29605

Query Match.	19.7%;	Score 340.4;	DB 11;	Length 471;
Best Local Similarity	96.8%;	Pred. No. 4.4e-88;		
Matches 358;	Conservative 0;	Mismatches 11;	Indels 1;	Gaps 1;
QY	1348	ATCCAGAGGACCTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGCTATGTCCAGAGACCTG	1407	
DB	55	ATCGAGAGGACCTTCATGCTTAAGCTTCTCCAGAGACCCCAATTCGTATGTCCAGATCTG	114	
QY	1408	CTCCGCTCCAGAGCCGGGACCTCAAGGTGATGACAGATGAGCCGGCAACCTCTGAAGAG	1467	
DB	115	CTCCGCTCCAGAGCCGGGACCTCAAAGGGATGACAGATGAGCCGGCAACCTCTGAAGAG	174	
QY	1468	GAGCGCCGGCTGAGTTCTACCAAGACCCCTGGTCCAGAGAGGCCGTCAAGTCGCTACTTC	1527	
DB	175	GAGCGCCGGCTGAGATCTACCAAGCCCTGGTCCAGAGAGGCCGTCAAGTCGCTACTTC	234	
QY	1528	TACTGCAAGATCCAGACGCGCAGGCAGGAGCTGGAGCAGTCGCTGGTTGTCGCAACACC	1587	
DB	235	TACTGCAAGATCCAGACGCGCAGGCAGGAGCTGGAGCAGTCGCTGGTTGTCGCAACACC	294	
QY	1588	TAGAGCCCAAAACAGCAGCAGCAGGAACTTTCAGCCGTGTCCCGGGCCCCAGCATTT	1647	
DB	295	TAAAGAGCCCAAAATAAGCAGCAGCAGGAACTTTCAGCCGTGTCCCGGGCCCCAGCATTT	354	
QY	1648	TTGCCCCGGGCTCCAGC-TCACTCCTCTGCCACCTTGGGGGTGTGGGGCTGGATTAAGAAGT	1706	
DB	355	TTGCCCCGGGCTCCAGCATCACTCTCTGCCACCTTGGGGGTGTGGGGCTGGATTAAGAAGT	414	
QY	1707	CATTCACTCTG	1716	
DB	415	CATTCTGCTG	424	

```

RESULT 7
US-09-960-352-12390
; Sequence 12390, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathaleagan, Negappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12390
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 53-LIB3058-001-Q1-K1-F2
US-09-960-352-12390

```

	Query Match	12.0%	Score 206.2;	DB 10;	Length 285;
	Best Local Similarity	83.0%;	Pred. No. 2.1e-49;		
	Matches 235;	Conservative 0;	Mismatches 48;	Indels 0;	Gaps 0;
QY	559	GTGACATCAGAGAGGCTCTGAACAGGCCCATGAAGCAAAAGCGGAAGCTGCGACTCTAT	618		
Db	3	GTGGTATTGCTGATGCTCTGAAGCTTTTACGAAGCAGAGCCAGCTGGGTGCGTAT	62		
QY	619	ATCTCCAACTTTTAAACCTGCGAAGTCTGATCTCAGGATTCGACGCGCAGCATTGCC	678		
Db	63	ATCTCCAACTTTTAAACCTGCTTAAGCCCGATGCTTAGGATTTCTGACGGCAGCATTGCC	122		
QY	679	TCCTGGAGCTACGGGTGGAGGGGAAGCTCTCTGATATCCACGAAACAGACGCGAAG	738		
Db	123	TCCTGGAGCTGTGGTGGAGGGGAAGCTCTCTGATACCCCGACTGGCATATCGGAAG	182		
QY	739	TTCTCTCTTTCTTCAAGAGTGTGGTCTATCGAGCTGGACAAAGATCTTTATGGCCCTGCAC	798		

Db 183 TTCTCGTCTTCATCAAGAGCTCGTCAATTGATCTGGACATAGACCTTTATGGCCCTGAC 242

Qy 799 AACCACTCTGTTGAGTGGCATCGGACACCCACGACCCAGGAGA 841

Db 243 AACCACTCTGTTGAGTGGCATCGGACACCCACGACCCAGGAGA 285

RESULT 8

US-09-879-536-796/c

; Sequence 796, Application US/09879536

; Patent No. US20020144298A1

; GENERAL INFORMATION:

; APPLICANT: Endege, wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; TITLE OF INVENTION: PRODUCTS

; FILE REFERENCE: CCD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/879,536

; CURRENT FILING DATE: 2001-09-21

; PRIOR APPLICATION NUMBER: US 60/088,801

; PRIOR FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: Fast-SEQ for Windows Version 3.0

; SEQ ID NO 796

; LENGTH: 452

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(452)

; OTHER INFORMATION: n = A,T,C or G

US-09-879-536-796

Query Match	11.9%	Score 205;	DB 10;	Length 452;			
Best Local Similarity	73.4%;	Pred. No. 5.6e-49;					
Matches	312;	Conservative 0;	Mismatches 90;	Indels 23; Gaps 3;			
Qy	480	GTGTGTC	CCGAGTCC	CAGGCTTACATG	GACCTCTTTGGCATTTTGAGAGGAAC	CTGGATCA	539
Db	424	GCAGGTAC	CAGAAATCC	CAGGCCATATG	GATCTCTTGGCTTTTGAAAGG	-A	366
Qy	540	AACATCAT	CGGAAGGG	GTGACATCC	CAGGAGGCTCTGAAGAGGCC	CATGAAGCAAAA	599
Db	365	GACTATCAT	CAGGAAC	-GGCTAGATAT	CCAAGAGGCCCTTGAAC	CGTCCCATCAAGCAAAA	307
Qy	600	CGGAAGCT	CGGACTCT	ATATCTTCAAC	ACTTTTAAACCTTCGGAAGTCT	GATGCTGAGGA	659
Db	306	ACGGAAGCT	CGGAATTTT	TCTATTTCTAAC	CTTTCAATCCGGCTAA	GTGATGCCGAGGA	247
Qy	660	TTCCGAGC	GGCAGCAT	TGCCTCCTCGGAGCT	ACGGGTGAGGGGAAGCT	CTCTGG	712
Db	246	TGGGGAAG	GGCAGCGT	TGCTTCTCGGAGCT	TTCGGGTAGAAAGC	GGCTCCTCGGAGGATC	187
Qy	713	-----	ATGATCC	CAGCAAA	CAGAGCGGAAGTCTCTCTTCTTCTTCAAGAG	758	
Db	186	AGCCTT	GTCAAATATAT	GTCGCCA	CTAAACAAAAGAGGAAGTCTCTCTCTTTTAAATC	127	
Qy	759	TTTGGT	CATCGAGCT	TGCACAAAGATCTTT	ATGGCCCTTGACAA	CCACCTCGTTGAGTGCGCA	818
Db	126	CTTGGT	GATTTGA	ACTTGACA	AGACCTGTATGGGCCAGACAA	CCATCTGGTAGAATGGCA	67
Qy	819	TCGGAC	ACCCACGAC	CC	CAGGAGACGGA	CGGCTTCCAGGTGAAACGGGCTCGGGACCTCGAG	878

Db 66 CAGAGCGGACCTACCGAGAGACCGATGCTTTCAGGTGAAGCGCGGAGACCTGAA 7
QY 879 TGTGC 883
Db 6 TGTAC 2

RESULT 9
US-09-918-995-2832

/ Sequence 2832, Application US/09918995
/ Publication No. US20030073623A1
/ GENERAL INFORMATION:
/ APPLICANT: Hyseq, Inc.
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
/ FILE REFERENCE: 20411-756
/ CURRENT APPLICATION NUMBER: US/09/918,995
/ PRIOR FILING DATE: 2001-07-30
/ PRIOR APPLICATION NUMBER: US/09/235,076
/ PRIOR FILING DATE: 1999-01-20
/ NUMBER OF SEQ ID NOS: 38054
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2832
/ LENGTH: 495
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(495)
/ OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2832

Query Match 11.8%; Score 203; DB 11; Length 495;
Best Local Similarity 75.6%; Pred. No. 2,2e-48;
Matches 251; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 662 CCGAGCGGACGATTCCTGCTGGAGCTACGGGTGAGGGGAGCTCTGGATGCCA 721
Db 164 CAGGGGACAAAGTGGCTCTCGGAACTCGAGTGAAGAAAGCTGGATGATCTTA 223
QY 722 GCAACAGAGCGGAAGTCTCTTCTTCTTCAAGATTGTGATCGAGCTGGACAAAG 781
Db 224 GCAACAGAGGAAGGAGTTTCTTCTTCTTAAAGCCTCGATGAGCTGGACAAAG 283
QY 782 ATCTTATGCGCCTGACCAACACCTGTTGAGTGGATGGACACCCAGACGAGAGA 841
Db 284 AGCTGTACGGGCTGACCAATCACTGTGAGTGGACCGGATGCCACACCCAGAGAGA 343
QY 842 CCGAGCGGCTTCCAGGTGAAGAGCGGCTGGGAGCTGAGTGGCGCTGACGCTGCTCTTA 901
Db 344 CAGATGCTTCCAAAGTAAAGCGGCTGGAGACCTCAACGTCAAGTCAACCTCTCTCTTA 403
QY 902 TGTGTGACTACGAGCTCCCAATTAAGTGGATCCCGGCTAGCGCGGCTGCTGGGC 961
Db 404 TGTGTGATCATCGAGCTCCCGCAATTAAGTGGATCCCGGCTAGCGCGGCTGCTGGAG 463
QY 962 TGCACACAGAGCGGCTGACGCTGATGTCAG 993
Db 464 TGCACACGAGAGCAAGGCGGCAATCATGACG 495

RESULT 10

US-09-918-995-34033
/ Sequence 34033, Application US/09918995
/ Publication No. US20030073623A1
/ GENERAL INFORMATION:
/ APPLICANT: Hyseq, Inc.
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
/ FILE REFERENCE: 20411-756
/ CURRENT APPLICATION NUMBER: US/09/918,995
/ CURRENT FILING DATE: 2001-07-30

/ PRIOR APPLICATION NUMBER: US/09/235,076
/ PRIOR FILING DATE: 1999-01-20
/ NUMBER OF SEQ ID NOS: 38054
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 34033
/ LENGTH: 420
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-918-995-34033

Query Match 10.5%; Score 180.4; DB 11; Length 420;
Best Local Similarity 94.4%; Pred. No. 7e-42;
Matches 187; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 903 GCTGACACTACGAGCTCCCAAGTTCAAACTGATCCCGGCTAGCGCGGCTGCTGGGCT 962
Db 3 GCTGACACTACGAGCTCCCAAGTTCAAACTGATCCCGGCTAGCGCGGCTGCTGGGCT 62
QY 963 GCACACAGAGCGGCTGAGCCTGATGCTCCAGGCGCTGTGGCAGTATGTGAAGACCAACAG 1022
Db 63 GCACACAGAGCGGCTGAGCCTGATGCTCCAGGCGCTGTGGCAGTATGTGAAGACCAACAG 122
QY 1023 GCTGACAGACTCCATGACACAGGAATACATCATGATGAGGACAAAGTATTTCCACAGATTTT 1082
Db 123 GCTGACAGACTCCATGACACAGGAATACATCATGATGAGGACAAAGTATTTCCACAGATTTCC 182
QY 1083 TGATGTCCCGGCTGAA 1100
Db 183 TGATGACAGAGCAGTA 200

RESULT 11

US-09-879-536-645/C
/ Sequence 645, Application US/09879536
/ Patent No. US20020144298A1
/ GENERAL INFORMATION:
/ APPLICANT: Endege, Wilson O.
/ APPLICANT: Steinmann, Kathleen E.
/ APPLICANT: Astle, Jon H.
/ APPLICANT: Burgess, Christopher C.
/ APPLICANT: Bushnell, Steven E.
/ APPLICANT: Carroll III, Eddie
/ APPLICANT: Carino, Theodore J.
/ APPLICANT: Derli, Adnan
/ APPLICANT: Ford, Donna M.
/ APPLICANT: Lewis, Marcia E.
/ APPLICANT: Monahan, John E.
/ APPLICANT: Schlegel, Robert
/ TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
/ FILE REFERENCE: CCD-257 (US)
/ CURRENT APPLICATION NUMBER: US/09/879,536
/ PRIOR FILING DATE: 2001-09-21
/ PRIOR APPLICATION NUMBER: US 60/088,801
/ PRIOR FILING DATE: 1998-06-10
/ NUMBER OF SEQ ID NOS: 850
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 645
/ LENGTH: 690
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(690)
/ OTHER INFORMATION: n = A,T,C or G
US-09-879-536-645

Query Match 10.4%; Score 179.2; DB 10; Length 690;
Best Local Similarity 64.5%; Pred. No. 1.8e-41;
Matches 323; Conservative 0; Mismatches 169; Indels 9; Gaps 4;

QY 1112 TTCCGAGGCGCTCAGAGCGGCTGATTTGCCCCCTGACCAATTTGATCA-ACCATGTC 1170

Db 638 TCCCTAAGNGGTTCCATNCNTGGTTAATGCGCCNCCANAAACCTATTNTCNTATTATCAATGTC 579
Qy 1171 ATCAGCGTGGACCTTTAGACAGAGAGAA-GACAGCGTGCTATGACATTTGACGTGGAGGT 1229
Db 578 ATCAGTGTGACCCCAATGTTNAGAAAAAGCCAGCTTGTATGCGCATTTGATGTTNAAAG 519
Qy 1230 GGAGGAGCCATTAAAGGGGAGATGAGCAGCTTCCT-----CCTATCCAGCGGCCAACCA 1283
Db 518 TGAATGCCNCCTTGAAGGCCCAAGATGATTTCTTTTCTGTGTCTCCANTGCCGGGCA 459
Qy 1284 GAGAGATCAGTCTCTGGAGAGTAAGATCCATGAGAGATTAGT-CCATAAACAGC 1342
Db 458 CCAGAGATTTGGTACTTTAGCCACCAAGATCCATGAGCCATAGAGAACCCATCAACCCAGC 399
Qy 1343 TCAAGATCCAGAGGACTTCATGCTAAAGCTTCTCCAGAGACCCCAAGAGCTATGTCCAAAG 1402
Db 398 TGAAGACTCAGCGGAGTTTCATGTTAAGTTTGGCCAGAGACCTCAGGGTTTCATCATG 339
Qy 1403 ACCTGCTCCGCTCCAGAGCGGGACCTCAAGGTGATGACAGATGACCGCGCAACCCCTG 1462
Db 338 ACTGGCTTCAGTCCCGAGTCAGGGACCTCAAGACAATGACTGATGTGGTGGTAACCCAG 279
Qy 1463 AAGAGAGCGCGGCTGAGTTCTACCAAGCCCTGTGCCAGAGCGCGTCAGTCGCT 1522
Db 278 AGGAGAGCGCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGGAGCTGTGCGCGAT 219
Qy 1523 ACTTCTACTGCAAGATCCAGAGCGCGAGAGCTGGAGCAGTCGCTGGTTGTGCGCA 1582
Db 218 ACTTCTACTCNAAGTGCAGAGACGACAGAGATTAGAGCAGCCCTGGGATCCGGA 159
Qy 1583 ACACCTAGGAGCCCAAAACA 1603
Db 158 ATACATAGGGCTCTCCACACA 138

RESULT 12

US-09-764-877-2295/c
; Sequence 2295, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2295
; LENGTH: 8788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2295

Query Match 9.9%; Score 170.4; DB 10; Length 8788;
Best Local Similarity 98.9%; Pred. No. 1.5e-38;
Matches 182; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1534 AAGATCCAGCAGCGGAGAGCTGGAGCAGTGTGCTGGTTCGCGCAACCTAGGAG 1593
Db 6674 AAGATCCAGCAGCGGAGAGCTGGAGCAGTGTGCTGGTTCGCGCAACCTAGGAG 6615
Qy 1594 CCCAAAAACAAGCAGCAGCGAACTTTTCAGCCGCTGTCGGGGCCCCCAGCATTTTGCCC 1653
Db 6614 CCCAAAAATAGCAGCAGCAGCGAACTTTTCAGCCGCTGTCGGGGCCCCCAGCATTTTGCCC 6555
Qy 1654 CGGGCTCCAGC-TCACCTCTTCGCCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCA 1712
Db 6554 CGGGCTCCAGCATCACTCTCTGCCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCA 6495
Qy 1713 TCTG 1716
Db 6494 TCTG 6491

RESULT 13

US-09-860-670-247/c
; Sequence 247, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL27P1
; CURRENT APPLICATION NUMBER: US/09/860,670
; PRIOR FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 8788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-860-670-247

Query Match 9.9%; Score 170.4; DB 10; Length 8788;
Best Local Similarity 98.9%; Pred. No. 1.5e-38;
Matches 182; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1534 AAGATCCAGCAGCGGAGAGCTGGAGCAGTGTGCTGGTTCGCGCAACCTAGGAG 1593
Db 6674 AAGATCCAGCAGCGGAGAGCTGGAGCAGTGTGCTGGTTCGCGCAACCTAGGAG 6615
Qy 1594 CCCAAAAACAAGCAGCAGCGAACTTTTCAGCCGCTGTCGGGGCCCCCAGCATTTTGCCC 1653
Db 6614 CCCAAAAATAGCAGCAGCAGCGAACTTTTCAGCCGCTGTCGGGGCCCCCAGCATTTTGCCC 6555
Qy 1654 CGGGCTCCAGC-TCACCTCTTCGCCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCA 1712
Db 6554 CGGGCTCCAGCATCACTCTCTGCCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCA 6495
Qy 1713 TCTG 1716
Db 6494 TCTG 6491

RESULT 14

US-09-918-995-27234
; Sequence 27234, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27234
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(440)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27234

Query Match 9.4%; Score 162; DB 11; Length 440;
Best Local Similarity 97.1%; Pred. No. 1.5e-36;
Matches 165; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 910 TACCAGCCTCCCGAGTTCAAACTGGATCCCGCTAGCCCGCTGCTGGGGCTGCACACA 969
Db |

Db 47 TGCCAGCTCCCGACTGTAATGATCCCGCTAGTTGGCTGGGGCTGCACACA 106
QY 970 CAGAGCGGCTCAGCCATTTCTCCAGGCGCTGTGCACTATGTGAAGCAACAGGCTGCAG 1029
Db 107 CAGAGCGGCTCAGCCATTTCTCCAGGCGCTGTGCACTATGTGAAGCAACAGGCTGCAG 166
QY 1030 GACTCCCATGACAGGAATATCATCATGGGAGCAAGTATTTCCAGCAGAT 1079
Db 167 GACTCCCATGACAGGAATATCATCATGGGAGCAAGTATTTCCAGCAGAT 216

RESULT 15

US-10-070-676-23/C
Sequence 23, Application US/10070676
Publication No. US20030059788A1

GENERAL INFORMATION:

APPLICANT: Toque, Bruno
APPLICANT: Bracco, Laurent
APPLICANT: Schweighoffer, Fabien
TITLE OF INVENTION: Genetic Markers of Toxicity, Preparation
TITLE OF INVENTION: and Uses
FILE REFERENCE: 50146/003002
CURRENT APPLICATION NUMBER: US/10/070,676
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: PCT/FR00/02503
PRIOR FILING DATE: 2000-09-12
PRIOR APPLICATION NUMBER: FR 99/11405
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: 09/456,370
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 310
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 3, 33, 34
OTHER INFORMATION: n = A,T,C or G
US-10-070-676-23

Query Match 8.0%; Score 137.8; DB 14; Length 310;

Best Local Similarity 69.1%; Pred. No. 1.3e-29; Mismatches 89; Indels 1; Gaps 1;

Matches 201; Conservative 0;

QY 991 CAGGCGCTGTGAGTATGTGAAGCAACAGGCTGCAGACTCCCATGACAGGAATAC 1050
Db 306 CAGGCGCTGTGAGTATGTGAAGCAACAGGCTGCAGACTCCCATGAGGCGGAGTAC 247
QY 1051 ATCAATGGGAGCAAGTATTTCCAGCAGATTTTGAATGTCCCGGCTGAAGTTTCTGAG 1110
Db 246 ATCAATGGGAGCAAGTATTTCCAGCAGATTTTGAATGTCCCGGCTGAAGTTTCTGAG 188
QY 1111 ATTCCCGGAGCGCTGCAGCGCTGATGTGCCCCGACCCATTTGTCATCAACCATGTC 1170
Db 187 ATTCCCGGAGCGCTGCAGCGCTGATGTGCCCCGACCCATTTGTCATCAACCATGTC 128
QY 1171 ATCAGCGTGAACCTTCAGACCGAGAGAGACAGCGTGTATGACATGAGCGTGAAGTG 1230
Db 127 ATTAGTGTGACCTTAACGACGAGAGAGACAGCGTGTATGACATGAGCGTGAAGTG 68
QY 1231 GAGGAGCCATTAAAGGGGAGATGAGCAGCTTCTCTATCCAGGCGCAAC 1281
Db 67 GACGAGCCACTGAAGGCCCAATGAGCAATTTTNNGGCTCTACACCATC 17

Search completed: October 11, 2003, 12:32:45
Job time : 358 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 09:10:51 ; Search time 2542 Seconds
(without alignments)
16483.445 Million cell updates/sec

Title: U66619
Perfect score: 1724
Sequence: 1 GAATTCGGCGCGAGCGGCC.....GTCAATTCATCTGGAAATTC 1724

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: EST:*
2: em_estba:*
3: em_estbm:*
4: em_estcm:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_estc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estfun:*
16: em_estcom:*
17: em_esthum:*
18: em_estinv:*
19: em_estpin:*
20: em_estvrt:*
21: em_estfun:*
22: em_estmam:*
23: em_estmus:*
24: em_estpro:*
25: em_estrod:*
26: em_estphg:*
27: em_estvrt1:*
28: gb_estc1:*
29: gb_estc2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1183.4	68.6	1669	AK005094	AK005094 Mus muscu
2	1016	58.9	1201	AL571327	AL571327 AL571327
3	965.6	56.0	1201	AL565228	AL565228 AL565228
4	796.8	46.2	1028	BMS52001	BMS52001 AGENCOURT

C	5	783.8	45.5	834	9	AL546132	AL546132
C	6	761	44.1	906	13	BX369145	BX369145
C	7	750.4	43.5	1050	12	B1524726	B1524726
C	8	741.8	43.0	851	12	B1524142	B1524142
C	9	740.6	43.0	904	13	B0720110	B0720110
C	10	730.2	42.4	842	12	B1553787	B1553787
C	11	728.8	42.3	807	12	B1520621	B1520621
C	12	724.6	42.0	958	12	B1409014	B1409014
C	13	698.4	40.5	1301	9	AL533695	AL533695
C	14	694.4	40.3	837	14	CD517671	CD517671
C	15	685.2	39.7	894	14	CD558853	CD558853
C	16	681.2	39.5	3206	11	BC000063	BC000063
C	17	679.6	39.4	1129	10	BF688178	BF688178
C	18	674.8	39.1	947	13	BQ955489	BQ955489
C	19	672.6	39.0	859	14	CA983665	CA983665
C	20	672.4	39.0	710	12	BM979567	BM979567
C	21	660.4	38.3	714	9	AI091285	AI091285
C	22	659.8	38.1	819	12	B1906319	B1906319
C	23	657.4	37.1	701	13	B0731539	B0731539
C	24	639.6	37.1	1163	12	BMS47645	BMS47645
C	25	634.2	36.8	704	13	BQ954817	BQ954817
C	26	631.6	36.6	911	13	B0670467	B0670467
C	27	624.6	36.2	674	9	AI423028	AI423028
C	28	624.2	36.2	732	14	CA323812	CA323812
C	29	616.6	35.8	805	11	AK019374	AK019374
C	30	616.4	35.8	863	12	B1601560	B1601560
C	31	615.6	35.7	815	12	BG825152	BG825152
C	32	614.6	35.6	717	9	AA902273	AA902273
C	33	612.6	35.5	811	14	B1712841	B1712841
C	34	610.2	35.4	813	10	BF972004	BF972004
C	35	607.8	35.3	729	12	B1598614	B1598614
C	36	607	35.2	615	9	AI300927	AI300927
C	37	596.8	34.6	805	10	BG170919	BG170919
C	38	595.8	34.6	961	10	BG115124	BG115124
C	39	595.2	34.5	877	13	B0521158	B0521158
C	40	592.2	34.4	936	13	B0591019	B0591019
C	41	586.8	34.0	947	13	BQ940444	BQ940444
C	42	582.2	33.8	664	9	AI564026	AI564026
C	43	581	33.7	628	14	CA449683	CA449683
C	44	576.4	33.4	603	12	BG819745	BG819745
C	45	569.8	33.1	679	10	BF722549	BF722549

ALIGNMENTS

RESULT 1
AK005094
LOCUS
DEFINITION Mus musculus adult male cerebellum cDNA, RIKEN full-length enriched library, clone:15000010J14 product:SMI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3, full insert sequence.
ACCESSION AK005094
VERSION AK005094.1 GI:12836786
KEYWORDS HTC: CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Carinci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL MEDLINE
PUBMED 99279253
10349636
2
Carinci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE
20499374

PUBMED
REFERENCE
AUTHORS

11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.P., Brownstein, M.J., Bult, C.,
Flatcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazza, R., Mombert, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5

PUBMED
REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1669)

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel.81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTT 3'], cDNA was

prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGCGCGCAATTAATTCGAGCTTAATAAATTAATCCCGCCCC 3']. cDNA
was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
end: SstI. Host: SOLR.

FEATURES

source

1. 1669
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1500001J14"
/db_xref="MGI:1895271"
/db_xref="taxon:10090"
/clone="1500001J14"
/sex="male"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
398. 1543
/dev_stage="adult"

CDS

/note="unnamed protein product; SWI/SNF related, matrix
associated, actin dependent regulator of chromatin,
subfamily d, member 3 (MGDI|MGI:1914243)
putative"

/codon_start=1
/protein_id="BAB23813.1"
/db_xref="GI:12836787"
/db_xref="MGI:1914243"

/translation="MADKLIPQIRILVPSQAYMDLLAFERKLDQTIMKRVDIQIA
LKPKQKQKRLRYISNTFNPQKDAEDSGSIASWELRVEGLDDPSKQKRFSSF
FKLSVLDELKDLGPDNHLVEMHRTPTQETDGFQVKKRPGDLVRCVTLMLLDYQPPQ
FKLDPRLRLGHTQSRSAIQVALWQYVKNRLQSDHKYINGDKYFQIQIFDQPL
KFCEIPORLTALLPPDPIVINHVISVDSDDOKKTACDIDIVEEBPKGOMSSFLLS
TANQQLSALDSKIHTIESINOLKIQRDPMLSFSDPKGVQDLRLRSQSDRLKVMTD
VAGNPEERRAEFYHQFWSQEAVSRYFYKIQQRQLEQSLVVRNT"

polyA_signal

1651. 1656
/note="putative"

polyA_site

1669
/note="putative"

BASE COUNT 410 a 493 c 499 g 277 t

ORIGIN

Query Match 68.6%; Score 1183.4; DB 11; Length 1669;
Best Local Similarity 91.8%; Pred. No. 1.9e-260;
Matches 1294; Conservative 0; Mismatches 111; Indels 4; Gaps 4;

Qy 219 GCGCCCCGGATGCGCTGTGAGCCCGGATGCCCCACCA-GGGCGGCCCATGGCCCCCCC 277
Db 170 GCGCCCCGGATGCGCTGTGAGCCCGGATGCCCCACCA-GGGCGGCCCATGGCCCCCCC 229
Qy 278 GGGCTCCCGTACATGGCAGCCCGCGCGGACCGGCTGCGCCCGCGCGG-ATGGA 336
Db 230 GGGCTCCCGTACATGGCAGCCCGCGGACCGGCTGCGCCCGCGCGG-ATGGA 289
Qy 337 GCGCCCGCGCAAGCGAGCAGCGCCCGCCCGCC-GGCAGAGCCAGGCACAGAGGCCA 395
Db 290 GCGCCCGCGCAAGCGAGCAGCGCCCGCCCGCC-GGCAGAGCCAGGCACAGAGGCCA 349
Qy 396 GCGGAGCCCGCCCGCCCGCGGAGCGCGAGTGCCTGCGAGGAGGAGTGGCTGACAA 455
Db 350 GCGCGTCCCA-CGCCCCAGCGGAGCGCGAGTGCCTGCGAGGAGGAGTGGCTGACAA 408
Qy 456 AATCCTCCCTCAAGAGGATTCGGGAGCTGGTCCCGAGTCCCGAGCTTACATGGACCTCTT 515
Db 409 AATCCTCCCTCAAGAGGATTCGGGAGCTGGTCCCGAGTCCCGAGCTTACATGGACCTCTT 468
Qy 516 GGCAATTTGAGAGGAACTGGATCAAAACCATCATCGCGAAGGGGTGGACATCCAGGAGGC 575
Db 469 AGCAATTTGAGAGGAACTGGATCAAAACCATCATCGCGAAGGGGTGGACATCCAGGAGGC 528
Qy 576 TCTGAAGAGGCCCATGAAGCAAAAGCGGAGCTCGGACTCTATATCTTCAACACTTTAA 635
Db 529 CTTGAAGAGGCCCATGAAGCAAAAGCGGAGCTCGGACTCTATATCTTCAACACTTTAA 588


```

QY 636 CCCTGCGAAGTCTGATGCTGAGGATTCGACGGCAGCATTCCTCTGGAGCTACGGGT 695
    |||||
Db 589 CCTCGGAAAGCTGATGCGGAAACCTCTATGGCAGCATTCCTCTGGAGCTGCGGT 648
QY 696 GGAAGGGAAGCTCTGGATGATCCGCAAAACGAACGGAAGTCTCTTTCTTTCAA 755
    |||||
Db 649 GGAGGGGAAGCTCTGGATGATCTCTAGTAAGCAGAGAGGAAGTCTTTCTTTCAA 708
QY 756 GAGTTTGGTATCGAGCTGGACAAAGATCTTTATGGCCCTGACCAACCTCTGTTAGAG 815
    |||||
Db 709 GAGTTTGGTATCGAGCTGGACAAAGATCTTTATGGCCCTGACCAACCTCTGTTAGAG 768
QY 816 GCATCGACACCCAGCAGCAGGAGAGCGGCTTCCAGGTGAACGGCTTGGGAGCT 875
    |||||
Db 769 GCACCGGACACCCAGCAGCAGGAGAGCGGCTTCCAGGTGAACGAGCAGGAGCTT 828
QY 876 GAGTGTGCTGACGCTGCTCTCTCATGTGACCTACAGCCCTCCAGTTCAACTGGA 935
    |||||
Db 829 GAGTGTGCTGACGCTGCTCTCTCATGTGACCTACAGCCCTCCAGTTCAACTGGA 888
QY 936 TCCCCCCTAGCCCGGCTGCTGGGGCTGACACACAGAGCCGCTACGCCATTTGTCAGGC 995
    |||||
Db 889 CCCCCCTTAGCCCGGCTGCTGGGGTTCACACACAGAGCCGCTACGCCATTTGTCAGGC 948
QY 996 CCGTGGCAGTATGTGAAGACCAACAGGCTGACGAGCTCCATGACAAAGAAATACATCA 1055
    |||||
Db 949 ACTGTGGCAGTATGTGAAGACCAACAGGCTGACGAGCTCCATGACAAAGAAATACATCA 1008
QY 1056 TGGGGCAAGATATTTTCAGCAGATTTTGTATGTCCCGGCTGAAGTTTCTGAGATTC 1115
    |||||
Db 1009 TGGCGCAAGATATTTTCAGCAGATTTTGTATGTCCCGGCTGAAGTTTCTGAGATTC 1068
QY 1116 CCAAGCCTTCAACAGCCTTGTATTTGCCCTGACCCCAATTGTATTAACATGTCTACG 1175
    |||||
Db 1069 CCAAGCCTTCAACAGCCTTGTATTTGCCCTGACCCCAATTGTATTAACATGTCTACG 1128
QY 1176 CGTGGACCTTGAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1235
    |||||
Db 1129 CGTGGACCTTGAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188
QY 1236 GCCATTAAAGGGGAGAGATGAGCAGCTTCTCTCTATTCACGGCCACACAGAGAGATCAG 1295
    |||||
Db 1189 ACCGCTGAAAGGGGAGAGATGAGTATGCTTCTCTCTGTCACGGCCACACAGAGAGATCAG 1248
QY 1296 TCTCTGACAGTATGATCCATGAGACGATTTGATTCATTAACAGCTCAAGTCCAGAG 1355
    |||||
Db 1249 TCTCTGACAGTATGATCCATGAGACGATTTGATTCATTAACAGCTCAAGTCCAGAG 1308
QY 1356 GGACTTCAATGCTAAGCTTCTCCAGAGACCCCAAGGCTATGTCAGAGCTCTGTCGCTC 1415
    |||||
Db 1309 GGACTTCAATGCTAAGCTTCTCCAGAGACCCCAAGGCTATGTCAGAGCTCTGTCGCTC 1368
QY 1416 CCAAGACCCGGAAGCTTCAAGGTGATGACAGATGATGAGCGGCAACCTGAAGAGAGCGCC 1475
    |||||
Db 1369 CCAAGACCCGGAAGCTTCAAGGTGATGACAGATGATGAGCGGCAACCTGAAGAGAGCGCG 1428
QY 1476 GAGTGTGCTGACGCTGCTCTCTCATGTGACCTACAGCCCTCCAGTTCAACTGGA 1535
    |||||
Db 1429 GAGTGTGCTGACGCTGCTCTCTCATGTGACCTACAGCCCTCCAGTTCAACTGGA 1488
QY 1536 GATTCGACGAGCGCAGCAGAGAGAGTGGAGAGTGGCTGTTGGCGGAACACCTAGAGAGCC 1595
    |||||
Db 1489 GATTCGACGAGCGCAGCAGAGAGTGGAGAGTGGCTGTTGGCGGAACACCTAGAGAGCC 1548
QY 1596 CAAAAACAAGCAGCAGCAGCACTTTCA 1624
    |||||
Db 1549 CGTGAACAAGCGCTCAGGCTGGAACAGCCA 1577

```

RESULT 2
AL571327/c
LOCUS AL571327 1201 bp mRNA linear EST 31-MAY-2003

DEFINITION AL571327 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1009Y109 3-PRIME, mRNA sequence.

ACCESSION AL571327

VERSION AL571327.2 GI:31292727

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Li, W.B., Gruber, C., Jessee, J., and Polayes, D. Full-length cDNA libraries and normalization Unpublished

AUTHORS On Feb 16, 2001 this sequence version replaced gi:12928512.

TITLE Contact: Genoscope

JOURNAL Genoscope - Centre National de Sequencage

COMMENT BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4359.r For more information about this cluster, see

http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1009AE05NP1&cluster=4359.r. Contact : Feng Liang Email : fliang@life-tech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0D1009AE05NP1.

FEATURES

source

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1009Y109"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 231 a 309 c 283 t 37 others

ORIGIN

Query Match 58.9%; Score 1016; DB 9; Length 1201;

Best Local Similarity 96.1%; Pred. No. 3,9e-222;

Matches 1097; Conservative 15; Mismatches 22; Indels 8; Gaps 7;

```

QY 528 GAACTGATCAAAACCATATCGGAGAGGGGTGACATCCAGAGGCTCTGAAGGCC 587
    |||||
Db 1161 GGAATGTGATCAAAACCATCAK--CGAAGCGGTGAGMMWYCCAGRRGCTCTGAAGAGCC 1104
QY 588 CATGAAGCAAAAGCGGAGAGCTGCTATATCTCCAAACCTTTAAACCTGGCAAGTC 647
    |||||
Db 1103 CATGAAGCAAAAGCGGAGAG-TCGACATATATTTCTCAACACTTTTAACCTT-CSAAGCC 1046
QY 648 TGATGTGAGATTCGACGAGAGATTCCTCTGAGAGCTTACGAGTGAAGGAGAGCT 707
    |||||
Db 1045 TGATGTGAGATTCGACGAGAGATTCCTCTGAGAGCTTACGAGTGAAGGAGAGCT 986
QY 708 CCGATGATGATC-CAGCAAAACGAAAGCGAAGTCTCTTTCTTTCTTCAAGAGTTGTCTA 766
    |||||
Db 985 CCGATGATGATCCCGCAGCAAAAGAGCGAAGTCTCTTTCTTTCTTCAAGAGTTGTCTA 926
QY 767 TCGACCT-GGAAGAGATCTTTATGCGCTGCAACACCTGTTGAGTGAAGTGGAGA 825
    |||||
Db 925 TCGAGCTGGGAAGAGATCTTTATGCGCTGCAACACCTGTTGAGTGAAGTGGAGA 866
QY 826 CCCAGACCCAGAGAGAGAGAGAGCTTCCAGTGAAGAGCGCTGGGAGCTGAGTGGCC 885
    |||||
Db 865 CCCAGACCCAGAGAGAGAGAGAGAGCTTCCAGTGAAGAGCGCTGGGAGCTGAGTGGT-SGC 807
QY 886 TCGACGCTGCTCTATGCTGAGTACAGCCTCCCAAGTTAAATGATATCCCGCCTTA 945
    |||||
Db 806 TCGACGCTGCTCTATGCTGAGTACAGCCTCCCAAGTTAAATGATATCCCGCCTTA 747

```


QY	1254	GAGCAGCTTCCTCCATATCAGCGCCAAACGACGAGAGATTCAGTCTCTCGACATTAAGAT	1313
Db	439	GAGCAGCTTCCTCCATATCAGCGCCAAACGACGAGATTCAGTCTCTCGACATTAAGAT	380
QY	1314	CCATGAGACGATTGAGTGCATATAACACGACTCAAGTCCAGAGGGACCTTCATGCTAAAGCTT	1373
Db	379	CCATGAGACGATTGAGTGCATATAACACGACTCAAGTCCAGAGGGACCTTCATGCTAAAGT	320
QY	1374	CTCCAGAGACCCCAAAAGGCTATGTCTCAAGACTCTGTCGCTCCAGAGCCGGACCTTCAA	1433
Db	319	CTCCAGAGACCCCAAAAGGCTATGTCTCAAGACTCTGTCGCTCCAGAGCCGGACCTTCAA	261
QY	1434	GATGATGACAGATGTAGAGCCGGCAACCCCTGAAGAGAGAGCGCGGCTGAGTTCTAACACCA	1493
Db	260	GATGATGACAGATGTAGAGCCGGCAACCCCTGAAGAGAGAGCGCGGCTGAGTTCTAACACCA	201
QY	1494	GCCCTGCTCCAGAGAGCCGCTAGTCTGCTACTTCTACTGCATAAGATCCAGACGCGCAGCA	1553
Db	200	GCCCTGCTCCAGAGAGCGCGCTAGTCTGCTACTTCTATTAATGCAAGATCCAGACGCGCAGCA	141
QY	1554	GGAGCTGAGAGAGTGCCTGCTGTTGGGCCAACACCTTAGAGCCCAAAAACAGGACGACG	1612
Db	140	GGAGCTGAGAGAGTGCCTGCTGTTGGTGGNNAACACTTAGAGAGCCCAAAAATTAAAGCACG	81
QY	1613	ACGGAACCTTTCAGCGCGTCTCCCGGGCCCGACGACATTTTGGC--CCGGGCTCCAGTCACTC	1670
Db	80	ACGGAACCTTTCAGCGCGTCTCCCGGGCCCGACGACATTTTGGCCTCCGGGANNCAAGAAACACTC	21
QY	1671	CTTCTGCCACCTTGGGGTGT	1689
Db	20	CTTCTGCCACCTTGGGGTGT	2

LOCUS	BM552001	1028 bp	mRNA	linear	EST 20-FEB-2002							
DEFINITION	AGENCOURT_6543453 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5549450 5', mRNA sequence.											
ACCESSION	BM552001											
VERSION	BM552001.1	GI:18789498										
KEYWORDS	EST.											
SOURCE	Homo sapiens (human)											
ORGANISM	Homo sapiens											
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.											
REFERENCE	1 (bases 1 to 1028)											
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/											
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)											
JOURNAL	Unpublished											
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LLM12260 row: a column: 03 High quality sequence stop: 640.											
FEATURES	Location/Qualifiers											
source	1..1028											

```

FEATURES
    source
        Location/Qualifiers
            1..1028
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5549450"
                /tissue_type="leiomyosarcoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_id="NH_MGC_71"
                /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 2.1 kb."

```

BASE COUNT	251 a	288 c	282 g	205 t	2 others
ORIGIN					
Query Match	46.2%; Score 796.8; DB 12; Length 1028;				
Beet Local Similarity	98.7%; Pred. No. 6,3e-112;				
Matches 834; Conservative	0; Mismatches 8; Indels 3; Gaps 3;				
QY	441	GAAATGAGCTGACAAAATCTCTCTCAAAGATTGCGGAGCTGATCCCCAGATGCCAGGC	500		
Db	1	GAAATGAGCTGACAAAATCTCTCTCAAAGATTGCGGAGCTGATCCCCAGATGCCAGGC	60		
QY	501	TTAATGAGACCTCTTGACATTTGAGAGAAACTGATCAACCATCTATGCGGAAGGGGGT	560		
Db	61	TTAATGAGACCTCTTGACATTTGAGAGAAACTGATCAACCATCTATGCGGAAGGGGGT	120		
QY	561	GGACATCCAGAGGCTCTGAAAGAGGCCATGAAGCAAAACCGAAGCTGAGCTCTATAT	620		
Db	121	GGACATCCAGAGGCTCTGAAAGAGGCCATGAAGCAAAACCGAAGCTGAGCTCTATAT	180		
QY	621	CTCCAAACCTTTTAACCTGCGAAGTGTGATGCTGAGGATTTCCAGCGGACATTTGCTTC	680		
Db	181	CTCCAAACCTTTTAACCTGCGAAGTGTGATGCTGAGGATTTCCAGCGGACATTTGCTTC	240		
QY	681	CTGGAGACTACGGGTGAGGGGAAAGCTCCCTGATGATCCCAAGCAACAGAACGGGAAGTT	740		
Db	241	CTGGAGACTACGGGTGAGGGGAAAGCTCCCTGATGATCCCAAGCAACAGAACGGGAAGTT	300		
QY	741	CTCTCTCTTTCTTCAAGAGTTTGGTCATCGAGCTGGAACAAGATCTTATAGGCTCTGACAA	800		
Db	301	CTCTCTCTTTCTTCAAGAGTTTGGTCATCGAGCTGGAACAAGATCTTATAGGCTCTGACAA	360		
QY	801	CCACCTCGTGAAGTGAATCGGACACCACGACCCAGAGAGCGGACGCTTCCAGTGA	860		
Db	361	CCACCTCGTGAAGTGAATCGGACACCACGACCCAGAGAGCGGACGCTTCCAGTGA	420		
QY	861	ACGGCTCTGGGAGACTGAGTGTGGCTGACAGGCTCTCTATGCTGGAATCAACAGCTTC	920		
Db	421	ACGGCTCTGGGAGACTGAGTGTGGCTGACAGGCTCTCTATGCTGGAATCAACAGCTTC	480		
QY	921	CCAGTTCAAACTGATCCCGGCTGAGCCGAGCTGCTGGGCTGACACACAGAGCCGCTC	980		
Db	481	CCAGTTCAAACTGATCCCGGCTGAGCCGAGCTGCTGGGCTGACACACAGAGCCGCTC	540		
QY	981	AGCCATTGTCAGGACCTGTGAGAGTGTGAAGACCAAGAGCTGAGAGCTCCATGA	1040		
Db	541	AGCCATTGTCAGGACCTGTGAGAGTGTGAAGACCAAGAGCTGAGAGCTCCATGA	600		
QY	1041	CAAGGAATTAATCAATGGGGAACAATTTTCCAGCAGATTTTGAATTTGTCCTCCGGCTGAA	1100		
Db	601	CAAGGAATTAATCAATGGGGAACAATTTTCCAGCAGATTTTGAATTTGTCCTCCGGCTGAA	660		
QY	1101	GTTTCTGAGATTTCCCGAGGCTCAAGGCTGCTATTTGCCCTGACCCCAATGTGCAT	1160		
Db	661	GTTTCTGAGATTTCCCGAGGCTCAAGGCTGCTATTTGCCCTGACCCCAATGTGCAT	720		
QY	1161	CAACCATGTCATCAGCTGAGACCTTGACAGCCAGAAAGACAGCGTGTATGACATTTGA	1220		
Db	721	CAACCATGTCATCAGCTGAGACCTTGACAGCCAGAAAGACAGCGTGTATGACATTTGA	780		
QY	1221	CGTGG-AGTGGAGAG-CCATTAAAGGGGCAATGACA-GCTTCTCTTATCCAGCGC	1277		
Db	781	CGTGGAGTGGAGAGACCCATTAAAGGGGCAATGAGAGGCTTCTCTTATCCAGCGC	840		
QY	1278	CAACG 1282			
Db	841	NCAAC 845			
RESULT 5					
AL546132/c	834 bp mRNA 11near EST 31-MAY-2003				
LOCUS	AL546132 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA				
DEFINITION	clone CSD01025Y102 3-PRIME, mRNA sequence.				

```

ACCESSION   AL546132
VERSION     AL546132.2  GI:31267966
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 834)
AUTHORS    Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12878951.
            Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 4359.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSODI025DF01NP1&cluster=4359.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSODI025DF01NP1.
            Location/Qualifiers
            1..834
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CSODI025YL02"
               /tissue_type="PLACENTA COT 25-NORMALIZED"
               /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
               /note="1st strand cDNA was primed with a NotI-oligo(dT)
               primer. Five prime end enriched, double-strand cDNA was
               digested with Not I and cloned into the Not I and EcoR V
               sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT  158 a 223 c 254 g 196 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
255
```

QY 399 GGAGCCACCGCCCGCGAGCCGACAGTGCAGAGAGAGAGATGCTGACAAAT 458
DB 1 GGTGCCACCGCCCGCGAGCCGACAGTGCAGAGAGAGAGATGCTGACAAAT 60
QY 459 CCTCCCTCAAGAGATTGCGAGAGTGTGCTCCGAGTCCAGGCTTACATGAGACCTTTGGC 518
DB 61 CTTCCCTCAAGAGATTGCGAGAGTGTGCTCCGAGTCCAGGCTTACATGAGACCTTTGGC 119
QY 519 ATTTGAGAGAACTGATTAACCATCATGCGGAAAGGGGTGACATCCAGAGGCTCT 578
DB 120 ATTTGAGAGAACTGATTAACCATCATGCGGAAAGGGGTGACATCCAGAGGCTCT 179
QY 579 GAAGAGGCCCATGATGAGAAAGGGAGGAGCTGAGCTTATATCTCCAGACTTTTAACC 638
DB 180 GAAGAGGCCCATGATGAGAAAGGGAGGAGCTGAGCTTATATCTCCAGACTTTTAACC 239
QY 639 TCGAAGTCTGATGCTGAGAGATTCGACGCGACATGCTCTCTGGAGCTACGGGTGA 698
DB 240 TCGAAGTCTGATGCTGAGAGATTCGACGCGACATGCTCTCTGGAGCTACGGGTGA 299
QY 699 GGGGAAGCTCTGAGATGATCCAGCAACAGAGCGAGATTCTTTCTTTCAAGAG 758
DB 300 GGGGAAGCTCTGAGATGATCCAGCAACAGAGCGAGATTCTTTCTTTCAAGAG 359
QY 759 TTTGTCATCGACTGAGCAAAAGATCTTTATGCGCTGACAAACACTCTGTTAGTGA 818
DB 360 TTTGTCATCGACTGAGCAAAAGATCTTTATGCGCTGACAAACACTCTGTTAGTGA 419
QY 819 TCGAAGCCACGACCCAGAGAGAGAGAGGCTTCAAGTGAAGGCTGCGGAGCTGAG 878
DB 420 TCGAAGCCACGACCCAGAGAGAGAGAGGCTTCAAGTGAAGGCTGCGGAGCTGAG 479
QY 879 TGTGCTGACGCTGCTCTCTCATGCTGACATCAAGCTCTCCAGTTCAACTGATCC 938
DB 480 TGTGCTGACGCTGCTCTCTCATGCTGACATCAAGCTCTCCAGTTCAACTGATCC 539
QY 939 CGGCTGAGCCCGGCTGCTGAGGCTGACACACAGAGCCGCTGACCATTTCCAGCCCT 998
DB 540 CGGCTGAGCCCGGCTGCTGAGGCTGACACACAGAGCCGCTGACCATTTCCAGCCCT 599
QY 999 GTGGCAGTATGTGAAGACCAACAGCTGACAGACTCCCATGACAGAGATACATCATG 1058
DB 600 GTGGCAGTATGTGAAGACCAACAGCTGACAGACTCCCATGACAGAGATACATCATG 658
QY 1059 GGAAGATATTTTCAGACAGATTTTGAATGTCTCCGCGCTGAAGTTTCTGAGATCCCA 1118
DB 659 GGAAGATATTTTCAGACAGATTTTGAATGTCTCCGCGCTGAAGTTTCTGAGATCCCA 718
QY 1119 GCGCTCTACAGCTGCTGCTCTCTCATGCTGACATCAAGCTCTCCAGTTCAACTGATCC 1177
DB 719 GAGCTCTACAGCTGCTGCTCTCTCATGCTGACATCAAGCTCTCCAGTTCAACTGATCC 778
QY 1178 TGGAACCTTCAGACAGAGAGAGAGAGCTGATGACATTTGAGTGAAGTGG--AGGA 1235
DB 779 GGAACCTTCAGACAGAGAGAGAGAGCTGATGACATTTGAGTGAAGTGG--AGGA 838
QY 1236 GCCATTAAAGGGCAGATGAGAGCTTCTCTATCCAGCGCCACACAGCAGAGATCAG 1295
DB 839 GCTTTAAAGGGCAGATGAGAGCTTCTCTATCCAGCGCCACACATGAGGAGTCCAC 898
QY 1296 TCTCTTGG 1303
DB 899 TGCTTTTG 906

RESULT 7
BIS24726/c 1050 bp mRNA linear EST 29-AUG-2001
LOCUS 60305155271 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201316 3'
DEFINITION mRNA sequence.
ACCESSION BIS24726
VERSION BIS24726.1 GI:15349518

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1050)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1504 row: 9 column: 13
High quality sequence start: 2
High quality sequence stop: 832.
Location/Qualifiers
1..1050
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5201316"
/lab_host="DH10B"
/clone_11b="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; site 1: NotI; site 2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleen. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC library."

BASE COUNT 216 a 283 c 316 g 235 t

ORIGIN
Query Match 43.5%; Score 750.4; DB 12; Length 1050;
Best Local Similarity 94.6%; Pred. No. 2.7e-161;
Matches 876; Conservative 0; Mismatches 36; Indels 14; Gaps 9;
QY 747 TTTCTTCAAGAGTTTGTCATGAGCTGACGACAAAGATCTTTATGCGCTGACACACC- 805
DB 934 TTTTTCAGAGATTGTCATGAGCTGACGACCTTGAACAAATCTTTATGCGCTGACGACACACC 875
QY 806 -TCGTTGAGTGCATCGACACCCACGA-----CCAGAGACGAGCG-CTTCCAGGTG 858
DB 874 TTGTTGAGATGTCATTCGGGACACCTCAGACCCCAAGAGAGAGCGGCTTCCAGGTG 815
QY 859 AAAAGCGCTGGGAGACCTGAGT-GTGGCTGACAGCGCTCTCATGCTGAGCTACAGGC 917
DB 814 AAAAGCGCTGGGAGACCTGAGTGTGCTGACAGCGCTCTCATGCTGAGCTACAGGC 755
QY 918 TCC-CCAGTTCAAACTGAT-CCCGGCTAGCCGGCT-GCTGGGCTCAGACAGAGAG 974
DB 754 TCTCTAGTTCAAACTGATTCCTCCGCTTAGCCCGGCTAGCGCTGAGGCTCAGACACAGAG 695
QY 975 CCGCTCAGCCATTGTCAGGCGCTGTCGAGTATGGAAGACCAACAGGCTGAGG-ACT 1033
DB 694 CCGCTCAGCCATTGTCAGGCGCTGTCGAGTATGGAAGACCAACAGGCTGAGGACT 635
QY 1034 CCCATGACAGAA--TACATCAATGGGGCAGATATTTCCAGAGATTTTGAATGTGCC 1092
DB 634 CCCATGACAGAAATTTACATGAGGGGCAAGTATTTCCAGAGATTTTGAATGTGCC 575
QY 1093 CGGCTGAAGTTTTCAGAGATTTCCCAAGCGCTCAGAGCCCTGATTTGCGCCCTGACCA 1152
DB 574 CGGCTGAAGTTTTCAGAGATTTCCCAAGCGCTCAGAGCCCTGATTTGCGCCCTGACCA 515

```
1153 ATTGTCATCAACCATGTTCATCAGCGTGACCCCTTCAGACCAAGAAAGACAGCGTGCTAT 1212
Db 514 ATTGTCATCAACCATGTTCATCAGCGTGACCCCTTCAGACCAAGAAAGACAGCGTGCTAT 455
Qy 1213 GACATTGACGTGGAGGTGGAGAGCCATTAAAGGGGCGAGATGAGCAGCTTCTCTATACC 1272
Db 454 GACATTGACGTGGAGGTGGAGAGCCATTAAAGGGGCGAGATGAGCAGCTTCTCTATACC 395
Qy 1273 ACGGCCAACACGACGAGATCAGTCTCTGACAGCTAGATCCATGACACGATTGAGTCC 1332
Db 394 ACGGCCAACACGACGAGATCAGTCTCTGACAGCTAGATCCATGACACGATTGAGTCC 335
Qy 1333 ATAAACCAAGCTCAAGATCCAGAGGACTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGC 1392
Db 334 ATAAACCAAGCTCAAGATCCAGAGGACTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGC 275
Qy 1393 TATGTCCAAAGCTGCTCCGCTCCAGAGCGGGACCTCAAGGTGATGACAGATGTAGCC 1452
Db 274 TATGTCCAAAGCTGCTCCGCTCCAGAGCGGGACCTCAAGGTGATGACAGATGTAGCC 215
Qy 1453 GGCACCTCTGAAGAGGAGCGCGGGCTGAGTCTACCAACGACCCCTGTCACGAGGACC 1512
Db 214 GGCACCTCTGAAGAGGAGCGCGGGCTGAGTCTACCAACGACCCCTGTCACGAGGACC 155
Qy 1513 GTCAAGTCTGCTACTTCTACTGCAAGATCCAGCAGCGCAGGAGCTGGAGCAGTCTGCTG 1572
Db 154 GTCAAGTCTGCTACTTCTACTGCAAGATCCAGCAGCGCAGGAGCTGGAGCAGTCTGCTG 95
Qy 1573 GTTGTGCGCAACACCTAGAGGCCCAAAACAGCAGCAGCGGAACTTTTCAGCGGTGC 1632
Db 94 GTTGTGCGCAACACCTAGAGGCCCAAAACAGCAGCAGCGGAACTTTTCAGCGGTGC 35
Qy 1633 CCGGGCCCCAGCATTTTCCCGCGGC 1658
Db 34 CCGGGCCCCAGCATTAACCGCCCGGTC 9

RESULT 8
BI524142
LOCUS 603051552F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201316 5',
DEFINITION mRNA sequence.
ACCESSION BI524142
VERSION BI524142.1 GI:15348934
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1504 row: 9 column: 13
High quality sequence stop: 850.
Location/Qualifiers
1. 851
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5201316"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
```

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH MGC Library."

BASE COUNT 208 a 243 c 231 g 168 t

ORIGIN

Query Match 43.0%; Score 741.8; DB 12; Length 851;
Best Local Similarity 95.6%; Pred. No. 2.4e-159;
Matches 816; Conservative 0; Mismatches 33; Indels 5; Gaps 5;

Qy 652 GCTGAGATTCCGACGCGCAGATTGCTCTCTGGG-AGCTACGGGTGGAGGAAAGCTCCT 710
Db 1 GCTGAGATTCCGACGCGCAGATTGCTCTCTGGGAGCTACGGGTGGAGGAAAGCTCCT 60
Qy 711 GGATGATCCCGACCAACAGAGCGGAAGTCTCTTTCTTCAAGAGTTTGGTCATCGA 770
Db 61 GGATGATCCCGACCAACAGAGCGGAAGTCTCTTTCTTCAAGAGTTTGGTCATCGA 120
Qy 771 GCTGCAAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTGGCATCGGACACCCAC 830
Db 121 GCTGCAAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTGGCATCGGACACCCAC 180
Qy 831 GACCCAGGACGACGCGCTTCAGGTGAAACGCGCTGGGGACCTGAGTGTGCGCTGCAC 890
Db 181 GACCCAGGACGACGCGCTTCAGGTGAAACGCGCTGGGGACCTGAGTGTGCGCTGCAC 240
Qy 891 GCTGCTCCTCATGCTGGACTACGAGCTCCCGAGTTCAAACCTGGATCCCGCTAGCCCG 950
Db 241 GCTGCTCCTCATGCTGGACTACGAGCTCCCGAGTTCAAACCTGGATCCCGCTAGCCCG 300
Qy 951 GCTGCTGGGCTGCACACAGAGCGCTCAGCCATTTGTCAGGCGCTGTGGCAGTATGT 1010
Db 301 GCTGCTGGGCTGCACACAGAGCGCTCAGCCATTTGTCAGGCGCTGTGGCAGTATGT 360
Qy 1011 GAAGACCAAGCGTGGAGCTCCCATGACAAAGAAATACATCAATGGGGAACAAGTATTT 1070
Db 361 GAAGACCAAGCGTGGAGCTCCCATGACAAAGAAATACATCAATGGGGAACAAGTATTT 420
Qy 1071 CCAGCAGATTTTGAATTTCCCGGCTGAAAGTTTCTGAGATTTCCCGAGCGCTCAGCAG 1130
Db 421 CCAGCAGATTTTGAATTTCCCGGCTGAAAGTTTCTGAGATTTCCCGAGCGCTCAGCAG 480
Qy 1131 CCTGCTATTGCCCCCTGACCCCAATTGTCAACCATGTTCATCAGCGTGGACCCCTTCAGA 1190
Db 481 CCTGCTATTGCCCCCTGACCCCAATTGTCAACCATGTTCATCAGCGTGGACCCCTTCAGA 540
Qy 1191 CCAGAAAGACAGCGCTGCTATGACATTGACGTTGAGGTGGAGGAGGACCATTTAAAGGGCA 1250
Db 541 CCAGAAAGACAGCGCTGCTATGACATTGACGTTGAGGTGGAGGAGGACCATTTAAAGGGCA 600
Qy 1251 GATGAGCAGCTTCTCTATCCAGGCAACACGAGCAGAGATCAGTCTCTGGACAGTAA 1310
Db 601 GATGAGCAGCTTCTCTATCCAGGCAACACGAGCAGAGATCAGTCTCTGGACAGTAA 660
Qy 1311 GATCCATGAGACGA-TTGAGTCCATAAACCCAGCTCAAGATCCAGAGGACCTTCATGCTAA 1369
Db 661 GATCCATGAGACGA-TTGAGTCCATAAACCCAGCTCAAGATCCAGAGGACCTTCATGCTAA 720
Qy 1370 GCTTCTCCAGAGACCCCAAGGCTATGTCGAAGACCTGCTCCGCTCCAGAGCGGGACC 1429
Db 721 AGCTTCTCCAGAGACCCCAAGGCTATGTCGAAGACCTGCTCCG-TCCAGAGCGGGCA-C 778
Qy 1430 TCAGGTTGATCAGATGTAGCCGCAACCTGAGAGGAGCGCGGCTGAGTCTTACC 1489
Db 779 TCAAGGTTGATCAGATGTAGCCGCAACCTGAGAGGAGCGCGGG-TGAGTTCTAAC 837

QY	1490	ACCAGCCCGTCC	1503	
Db	838	AACGCTGTGTC	851	
RESULT 9				
LOCUS	B0720110			
DEFINITION	B0720110	904 bp	mRNA	linear
ACCESSION	AGNC000001.1	149144	lupsk1	sympathetic_trunk Homo sapiens cDNA clone
VERSION	B0720110			
KEYWORDS	EST.	GI:21859007		
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 904)			
TITLE	NIH-MGC http://mgc.nci.nih.gov/.			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished			
	Contact: Robert Strausberg, Ph.D.			
	Email: cga@bbsr.nlm.nih.gov			
	Tissue Procurement: Dr. James R. Lupski			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:			
	http://image.llnl.gov			
	Plate: LLM13586	row: 0	column: 10	
	High quality sequence stop: 599.			
FEATURES	Location/Qualifiers			
source	1..904			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6189537"			
	/sex="male"			
	/tissue_type="sympathetic trunk"			
	/dev_stage="adult, 16 yr"			
	/lab_host="MDH10B"			
	/clone_lib="lupsk1_sympathetic_trunk"			
	/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1: NotI; Site_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCGAGCGGTCCG-3' and 5'-GACACTGTCTGATGTCGACGCGGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies"			
BASE COUNT	191 a	286 c	254 g	173 t
ORIGIN				
Query Match	43.0%;	Score 740.6;	DB 13;	Length 904;
Best Local Similarity	94.7%;	Pred. No. 4.6e-159;		
Matches	810;	Conservative 0;	Mismatches 39;	Indels 6;
			Gaps 4;	
QY	243	CCGATGCCCCACCAAGGCGCCCATGGGCCCCCGGGCTCTCCCGTACATGGCAGCCCC	302	
Db	1	CCGATGCCCCACCAAGGCGCCCATGGGCCCCCGGGCTCTCCCGTACATGGCAGCCCC	60	
QY	303	GCCGTGCGACCCGGCTGGCCCCCGGGGATGAGCCCGCCCGGAAGCGACAGCGCCCC	362	
Db	61	GCCGTGCGACCCGGCT--CCCGGGGCTATGAGCCCGCCCGGAAGCGACAGCGCCCC	118	
QY	363	CGCCC-GGCAGAGCCAGGACAGAGCCAGGCGCAGCCGAGCCACCGCCCCCGCGGGA	421	
Db	119	CGCCCGGCAAGCCAGGACAGAGCCAGGCGCAGCCGAGTCCACCGCCCCCGCGGGA	178	
QY	422	GCCCGAGTCCCAAGAGAGAGATGCTGACAAATCTCTCCCTCAAGGATTCGGGAGC	481	

Db	179	GC	CC	CA	GT	CC	CA	AG	GA	GA	GA	AT	GG	CT	GA	AA	AT	CT	CC	TA	AA	GG	AT	TC	GG	AG	CC	238	
Qy	482	TG	GT	CC	CC	CG	AT	TC	CC	AG	CT	TAC	TG	AC	CT	CT	TG	GC	AT	TT	GA	GA	GA	AA	CT	GA	TC	541	
Db	239	TG	GT	CC	CC	AG	AT	TC	CC	AG	CT	TAC	TG	AC	CT	CT	TG	GC	AT	TT	GA	GA	GA	AA	CT	GA	TC	298	
Qy	542	CC	AT	CA	TG	CC	GA	AG	GG	GG	TG	GC	AC	TC	CA	GA	AG	CT	CT	GA	AG	GG	CC	AT	GA	AG	CA	601	
Db	299	CC	AT	CA	TG	CC	GA	AG	GG	GG	TG	GC	AC	TC	CA	GA	AG	CT	CT	GA	AG	GG	CC	AT	GA	AG	CA	358	
Qy	602	GG	AG	CT	GC	CA	CT	TA	TAT	CT	CC	AA	GC	CT	TT	TA	CC	CT	GG	AG	CT	CG	AT	GC	TC	GA	GA	661	
Db	359	GG	AG	CT	GC	CA	CT	TA	TAT	CT	CC	AA	GC	CT	TT	TA	CC	CT	GG	AG	CT	CG	AT	GC	TC	GA	GA	418	
Qy	662	CC	GA	CG	CG	AG	CA	TG	CC	CT	CT	CG	TG	AG	CT	TA	CG	GG	GT	GA	GG	GA	AG	CT	CT	GA	AT	721	
Db	419	CC	GA	CG	CG	AG	CA	TG	CC	CT	CT	CG	TG	AG	CT	TA	CG	GG	GT	GA	GG	GA	AG	CT	CT	GA	AT	478	
Qy	722	G	C	A	A	A	C	G	A	A	C	G	G	A	G	T	T	C	T	T	C	T	T	C	T	T	T	C	781
Db	479	G	C	A	A	A	C	G	A	A	C	G	G	A	G	T	T	C	T	T	C	T	T	C	T	T	T	C	538
Qy	782	A	T	C	T	T	T	T	G	C	C	T	G	A	A	C	C	A	C	T	C	T	G	A	A	C	C	A	841
Db	539	A	T	C	T	T	T	T	G	C	C	T	G	A	A	C	C	A	C	T	C	T	G	A	A	C	C	A	598
Qy	842	C	G	A	C	G	G	C	T	T	C	A	G	T	G	A	A	C	G	G	C	T	G	G	A	C	C	T	901
Db	599	C	G	A	C	G	G	C	T	T	C	A	G	T	G	A	A	C	G	G	C	T	G	G	A	C	C	T	658
Qy	902	T	G	C	T	G	A	C	T	A	C	C	A	C	C	T	C	C	C	A	G	T	T	C	A	A	C	T	961
Db	659	T	G	C	T	G	A	C	T	A	C	C	A	C	C	T	C	C	C	A	G	T	T	C	A	A	C	T	718
Qy	962	T	G	C	A	C	A	C	A	G	A	G	C	C	G	T	C	A	G	C	A	T	T	T	C	A	G	C	1020
Db	719	T	G	C	A	C	A	G	A	G	C	C	G	T	C	A	G	C	A	T	T	T	C	A	G	C	C	A	778
Qy	1021	A	G	C	T	G	C	A	G	A	C	T	C	C	A	G	A	T	A	T	C	A	T	C	A	T	C	A	1078
Db</																													


```
http://image.llnl.gov
Plate: L1AM11660 row: d column: 10
High quality sequence stop: 831.
FEATURES
    source
        Location/Qualifiers
            1..842
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5261913"
                /tissue_type="hippocampus"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_95"
                /note="Organ: brain; Vector: pBluescriptR (modified
                pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
                ); Oligo-dT primed using primer 5'-TCTTTTCTTTTCTTTVN-3',
                size-selected for average insert size 2.5 kb and
                normalized to 200 ng. This is a primary library enriched
                for full-length clones and constructed using the
                Cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIMH/NIHRI, National
                Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      165 a   275 c   282 g   120 t
ORIGIN
Query Match      42.4%; Score 730.2; DB 12; Length 842;
Best Local Similarity 98.3%; Pred. No. 1.le-156;
Matches 801; Conservative 0; Mismatches 8; Indels 6; Gaps 6;
QY 17 GCCGAGCGGAGCGCGGAGCGGAGCGGCGCGCGCGCTCGGGCGCGGGTCCCGGGG 76
DB 29 GCGCGAGCGGAGCGGAGCGGAGCGGCGCGCGCGCTCGGGCGCGGGTCCCGGGG 88
QY 77 AGCAGATCTCAGAAATGCGCTTGCTGTGTCAGGCGCGGTGGCTCCGGGCGCGGAC 136
DB 89 AGCAGATCTCAGAAATGCGCTTGCTGTGTCAGGCGCGGTGGCTCCGGGCGCGGAC 148
QY 137 GAGGGGCACTGGATGACTCTCCAGGTGAGGACCTCCCATCTATGACTCCAGGTCTTC 196
DB 149 GAGGGGCACTGGATGACTCTCCAGGTGAGGACCTCCCATCTATGACTCCAGGTCTTC 208
QY 197 AGCACCCACCCAGTGTGTACAGCGCCCGGGATGCGCTCTGGAGCCCGGATGCCAC 256
DB 209 AGCACCCACCCAGTGTGTGTACAGCGCCCGGGATGCGCTCTGGAGCCCGGATGCCAC 268
QY 257 A-GGGGCGCCCATGCGCCCGCGCTCCCGTACATGGGAGCGAGCCCGCGCGGACCG 315
DB 269 AGGGGGCGCCCATGCGCCCGCGCTCCCGTACATGGGAGCGAGCCCGCGCGGACCG 328
QY 316 GCCTGGCGCCCGCGGG-ATGGAGCCCGCGCGCAAGCGAGCGCGCCCGCGCC-GGCAGA 373
DB 329 GCCTGGCGCCCGCGGGCATGGAGCCCGCGCGCAAGCGAGCGCGCCCGCGCGGCGAGA 388
QY 374 GCCAGGCAAGAGCCAGGCGCGGAGCGGAGCGGAGCGGCGCGCGGAGCGGAGTGCGCA 433
DB 389 GCCAGGCAAGAGCCAGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGTGCGCA 448
QY 434 AGAGGAGCAAGATGCTGCACAAAATCTCCCTCAAAGGATTCGGGAGTGTCTCCCGAGT 493
DB 449 AGAGGAGCAAGATGCTGCACAAAATCTCCCTCAAAGGATTCGGGAGTGTCTCCCGAGT 508
QY 494 CCCAGGCTTATCATGACCTCTTTGGCATTTGAGAGGAAACTGGATCAAAACCATCATCGGA 553
DB 509 CCCAGGCTTATCATGACCTCTTTGGCATTTGAGAGGAAACTGGATCAAAACCATCATCGGA 568
QY 554 AGGGGTTGGAATTCAGAGGCTCTGAAGAGGCGGATGAAGCAAAAGCGGAGTGGCGAC 613
DB 569 AGCGGTTGGAATTCAGAGGCTCTGAAGAGGCGGATGAAGCAAAAGCGGAGTGGCGAC 628
QY 614 TCTATATCTCAACACTTTTAACTCTGGAAGTCTGATGCTGAGGATTCGAGGCGGAGCA 673
DB 629 TCTATATCTCAACACTTTTAACTCTGGAAGTCTGATGCTGAGGATTCGAGGCGGAGCA 688
QY 674 TTGCTCTCTGGAGCTACGGGTGGAGGGAAGCTCTCTGGATGATCCCGAGCAACAGAGC 733
```

```
|||||
689 TTGCTCTCTGGAGCTACGGGTGGAGGGAAGCTCTCTGGATGATCCCGAGC-AAACAGAGC 747
|||||
734 GGAAGTTCTCTTCTTTCAAGAGTTTGGTCTATCGAGCTGGACAAAGATCTTTATGGCC 793
|||||
748 GGAAGTTCTCTTCTTTCAAGAG-TTGGTCTATCGAGCTGGACAAAGATCTTTATGGCC 806
|||||
794 CTGACACCACTCTGTTGAGTGGCATCGGACACC 828
|||||
807 CTGAC-AGCACCTCTGTTGAGTGGCATCGGAAACC 840
|||||

RESULT 11
BI520621/c
LOCUS      807 bp      mRNA      linear      EST 29-AUG-2001
DEFINITION 603071572T1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5163880 3',
            mRNA sequence.
ACCESSION  BI520621
VERSION    BI520621.1 GI:15345413
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 807)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-rc@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM11406 row: o column: 17
            High quality sequence start: 29
            High quality sequence stop: 805.
FEATURES
    source
        Location/Qualifiers
            1..807
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5163880"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_119"
                /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: EcoRV (destroyed); RNA source normal medulla from
                anonymous male age 27. Library is oligo-dT primed and
                directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.3 kb, insert size range
                0.9-3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research Genetics tracking code 013. Note:
                this is a NIH_MGC Library."
BASE COUNT      150 a   219 c   242 g   196 t
ORIGIN
Query Match      42.3%; Score 728.8; DB 12; Length 807;
Best Local Similarity 96.3%; Pred. No. 2.2e-156;
Matches 778; Conservative 0; Mismatches 27; Indels 3; Gaps 3;
QY 867 TGGGACCTGAGTGTGGCTGCGCTGCTCTCTATGCTGGACTACAGCTCCCGAGTT 926
DB 807 TGGGACCGGAGTG-GGCTGCGCTGCTCTCTATGCTGGACTACAGCTCCCGAGTT 749
|||||
927 CAAACTGGATCCCGCTAGCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCAT 986
|||||
748 CAAACTGGATCCCGCTAGCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCAT 689-
```


Db 611 CTCCTGTCCACGGTCAACAGCAGAGATCAGTGTCTGGACAGTAAGATCCATGAGACG 670
Qy 1324 ATTGAGTCCATAAACACAGCTCAAGATCCAGAGGAGTTCATGCTAAGCTTCTCCAGAGAC 1383
Db 671 ATTGAGTCCATAAACACAGCTCAAGATCCAGAGGAGTTCATGCTAAGTTCCTCCAGAGAC 730
Qy 1384 CCCAAGGCTATGTCCAAAGACCTGCTCCGCTCCCAAGAGCCGAGCCTCAAGGTG-ATGAC 1442
Db 731 CCCAAGGCTAGTCCAAAGACCTGCTCCGCTCCCAAGAGCCGAGTTCCTCAAGGTGAATGAC 790
Qy 1443 AGATGTAGCCGCAACCTGTAAGAGAGGCGC-CGGGCTGAGTTCACCAACCA-GCCCTGG 1500
Db 791 CGATGTGCGAGGAACCCCGAGGAAGAACGCGAGGGCTCAGTTCCTACCAACAGGTCTCTGG 850
Qy 1501 TCCAGGAGGCGCTGAG-TGCGTACTTCTACTGCAAGATCCAGCAGCC-AGGCAGGAGC 1558
Db 851 TCCAGGAGGCGTGTAGCCGCTACTTCTACTGTGAGATCCAGCAGCCAGGCGAGGAGC 910
Qy 1559 TGGAGCAGTCCGTGGTT-GTGGCAACACCTAGGAGCCC 1596
Db 911 TGAACAGTCCGTGGTTGTTGGCAACAGCTAGGAGCCC 949

RESULT 13
LOCUS AL533695 1201 bp mRNA linear EST 12-MAY-2003
DEFINITION AL533695 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF002YC19 5-PRIME, mRNA sequence.
ACCESSION AL533695
VERSION AL533695.2 GI:30538791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12797188.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4359.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF002AB10Qp1&cluster=4359.r. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF002AB10Qp1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF002YC19"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

FEATURES

source
1..1201
274 a 332 c 353 g 198 t 44 others
BASE COUNT 274 a 332 c 353 g 198 t 44 others
ORIGIN
Query Match 40.5%; Score 698.4; DB 9; Length 1201;
Best Local Similarity 96.0%; Pred. No. 2.4e-149;
Matches 789; Conservative 19; Mismatches 5; Indels 9; Gaps 9;

Qy 219 GCGCCCGGATGCGTCTGGAGCCCGATGCCCAACA-GGGCGGCCATGGGCCCCCC 277
Db 226 GCGCCCGGATGCGTCTGGAGCCCGATGCCCAACAAGGGGCGCCATGGGCCCCCC 285
Qy 278 GGGTCCCGTACATGGGAGCCCGCGCTGCGACCCGGCTGGCCCCCGGGG-ATGGA 336
Db 286 GGGTCCCGTACATGGGAGCCCGCGCTGCGACCCGGCTGGCCCCCGGGCATGGA 345
Qy 337 GCGCCCGGAAGCGAGCAGCCCGCCGCC- GGCAGAGCAGGACAGAGCCAGGGCCA 395
Db 346 GCGCCCGGAAGCGAGCAGCCCGCCGCCCGGCGAGCCAGGACAGAGCCAGGGCCA 405
Qy 396 GCGGAGCCCAACCGCCCGCGGAGCCCGAGTGCACAGAGGAGGAAGTGGCTGACAA 455
Db 406 GCGGAGCCCAACCGCCCGCGGAGCCCGAGTGCACAGAGGAGGAAGTGGCTGAWAA 464
Qy 456 AATCCTCCCTCAAAAGGATTCGGGAGTGGTCCCGAGTCCAGGCTTACATGAGACCTTT 515
Db 465 AATCCTCCCTCAAAAGGATTCGGGAGTGGTCCCGAGTCCAGGCTTACATGAGACCTTT 524
Qy 516 GGCATTTGAGAGGAACCTGGATCAAAACCATCATGCGAAGGGGTGGACATCCAGGAGGC 575
Db 525 GGCATTTGAGAGGAACCTGGATCAAAACCATCATGCGAAGGGGTGGACATCCAGGAGGC 584
Qy 576 TCTGAAGAGGCCCATGAAGCAAAAGCGGAAGCTCGCACTCTATATCTCCAACACTTTTAA 635
Db 585 TCTGAAGAGGCCCATGAAGCAAAAGCGGA- STGCACTCTATATCTCCAACACTTTTAA 643
Qy 636 CCCTGCGAAGTCTGATGCTGAGGATTCGACGGCAGCATTTGCCCTCTGGGAGCTACGGGT 695
Db 644 CCCTGCGAA-CCTGATGCTGAGGATTCGACGGCAGCATTTGCCCTCTGGGAGCTACGGGT 702
Qy 696 GGAGGGGAAGCTCCTGATGATCCAGCAAAACAGAGCGGAAGTCTCTCTCTTTCTTCAA 755
Db 703 GGAGGGGAAGCTCCTGATGATCCAGCAAAACAGAGCGGAAGTCTCTCTCTTTCTTCAA 762
Qy 756 GAGTTTGGTTCATCGAGCTGGCAAAAGATCTTTATGGCCCTGACAAACACCTCGTTGAGTG 815
Db 763 GAGTTTGGTTCATCGAGCTGGCAAAAGATCTTTATGGCCCTGACAAACACCTCGTTGAGTG 822
Qy 816 GCATCGACACCCACGACCCAGGAGACGCGCTTCAGGTGAAACGGCT- GGGGACC 874
Db 823 GCATCGACACCCACGACCCAGGAGACGCGCTTCAGGTGAAACGGCTGGGGACC 882
Qy 875 TGAGTGTGCGTGCACCGTCTCTCATGTGGACTACGAGCTCCCGAGTTCAAATGG 934
Db 883 TGAGTGTGCGTGCACCGTCTCTCATGTGGACTACGAGCTCCCGAGTTCAAATGG 942
Qy 935 ATCCCGCCTAGCCCGCTGCTGGGGCTGCACACAGAGCGGCTTCAGCCATTGTCAGG 994
Db 943 AT-CCCGCCTAGCCCGCTGCTGGGGTGCACACACAGAGCGGCTTCAGCCATTGTCAGG 1001
Qy 995 CCCTGTGGCAGTATGTCAAGACCAACAGGCTGCGAGGACTCCC 1036
Db 1002 CCCTGTGGCAGTATGTCAAGACCAACAGGCTGCGAGGACTCCC 1042

RESULT 14

LOCUS CD517671 837 bp mRNA linear EST 06-JUN-2003
DEFINITION AGENCOURT 14372899 NIH MGC 181 Homo sapiens cDNA clone
IMAGE:30400116 5', mRNA sequence.
ACCESSION CD517671
VERSION CD517671.1 GI:31449389
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 837)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
REFERENCE
AUTHORS
TITLE

JOURNAL COMMENT

Unpublished
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDM487 row: 1 column: 13
High quality sequence stop: 614.
Location/Qualifiers

FEATURES

source

```
1. 837
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:30400116"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH MGC 181"
/notes="Vector: PCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(invitrogen). Note: this is a NIH_MGC Library."
```

BASE COUNT

174 a 276 c 258 g 129 t

ORIGIN

Query Match 40.3%; Score 694.4; DB 14; Length 837;
Best Local Similarity 96.8%; Pred. No. 1.7e-148;
Matches 784; Conservative 0; Mismatches 16; Indels 10; Gaps 7;

218 AGGCGCCCGGAGTCCGTGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 276
1 AGGCGCCCGGAGTCCGTGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 60
277 CGGGCTCCCGTACATGGGAGCCCGCGTGGACCCGCGCTGGCCCGCGGG-ATGG 335
61 CGGGCTCCCGTACATGGGAGCCCGCGTGGACCCGCGCTGGCCCGCGGG-ATGG 120
336 AGCGCGCCCGGAGTCCGTGAGCCCGCGTGGACCCGCGCTGGCCCGCGGG-ATGG 394
121 AGCGCGCCCGGAGTCCGTGAGCCCGCGTGGACCCGCGCTGGCCCGCGGG-ATGG 180
395 AGCGCGCCCGGAGTCCGTGAGCCCGCGTGGACCCGCGCTGGCCCGCGGG-ATGG 454
181 AGCGCGCCCGGAGTCCGTGAGCCCGCGTGGACCCGCGCTGGCCCGCGGG-ATGG 240
455 AATCTCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCCGAGCTTACATGACCTCT 514
241 AATCTCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCCGAGCTTACATGACCTCT 300
515 TGGCATTGAGAGAACTGGATCAACATCATGCGGAAGGGGGTGAAGATTCAGAGAG 574
301 TGGCATTGAGAGAACTGGATCAACATCATGCGGAAGGGGGTGAAGATTCAGAGAG 360
575 CTCTGAAGAGCCCAAGCAAAAGCGAGAGTGGAGCTCTATATCTCCAACTTTTA 634
361 CTCTGAAGAGCCCAAGCAAAAGCGAGAGTGGAGCTCTATATCTCCAACTTTTA 420
635 ACCCTCGAAGTCTGATGCTGAGAGTTCGAGCGGAGCATGCTCTGAGAGCTAGCGG 694
421 ACCCTCGAAGTCTGATGCTGAGAGTTCGAGCGGAGCATGCTCTGAGAGCTAGCGG 480
695 TGGAGGGGAGGCTCTGATGATCCAGCAACAGAGCGGAGATTCTCTTTCTTCA 754
481 TGGAGGGGAGGCTCTGATGATCCAGCAACAGAGCGGAGATTCTCTTTCTTCA 540

QY 755 AGAGTTGGTCACTGAGCTGACAAAGATCTTATAGCCCTGACAAACCACTGTTGACT 814
DB 541 AGAGTTGGTCACTGAGCTGAGCAAAAGATCTTATAGCCCTGACAAACCACTGTTGACT 600
QY 815 GGCATGAGACACCAACGACCAAGAGAGAGAGAGGCTTCCAGGTAAAGGCTT-GGGAC 873
DB 601 GGCATGAGACACCAACGACCAAGAGAGAGAGAGGCTTCCAGGTAAAGGCTTGGGGGAC 660
QY 874 CTGAGTGGGCTGAGAGCTGCTCTCATGCTGAGCTACCAAGCTCCAGTTCAACATG 933
DB 661 CTGAGTGGGCTGAGAGCTGCTCTCATGCTGAGCTACCAAGCTCCAGTTCAACATG 720
QY 934 GATCCCGGCTGAGCCCGG--CTGCTGGGCTGACACACAGAGCCG--CTCAGCCATTGT 989
DB 721 GATCCCGGCTGAGCCCGGCTGCTGAGGCTGACACACAGAGCCGCTCAGCCCTTTGT 780
QY 990 CCAGGCCC--TGTGCGATGATGTAAGACC 1017
DB 781 CCAGGCCCCTGCTGCTGATGTAAGAAC 810

RESULT 15

CD558853 894 bp mRNA linear EST 11-JUN-2003
LOCUS AGENCOURT 14394871 NIH MGC 181 Homo sapiens CDNA clone
DEFINITION IMAGE:30400116 5', mRNA sequence.
ACCESSION CD558853
VERSION CD558853.1 GI:31584921
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 894)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDM487 row: 1 column: 13
High quality sequence stop: 615.
Location/Qualifiers

FEATURES

source

```
1. 894
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:30400116"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH MGC 181"
/notes="Vector: PCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(invitrogen). Note: this is a NIH_MGC Library."
```

BASE COUNT

195 a 289 c 270 g 139 t

ORIGIN

Query Match 39.7%; Score 685.2; DB 14; Length 894;
Best Local Similarity 96.2%; Pred. No. 2.3e-146;
Matches 744; Conservative 0; Mismatches 24; Indels 5; Gaps 4;

